

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 4, 2002, 08:38:05 ; Search time 4179.27 Seconds
(without alignments)
6481.608 Million cell updates/sec

Title: US-09-052-089a-7
Perfect score: 2007
Sequence: 1 GTGGCGTGCAGCAAAATTG.....AAAAAAAAAAAAAAAAAAAA 2007

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1075.2	53.6	2696	11	AK012948	AK012948 Mus muscu
2	1067.2	53.2	1938	11	AK012786	AK012786 Mus muscu
3	894.4	44.6	959	9	AL560947	AL560947 AL560947
4	863.4	43.0	887	9	AL560912	AL560912 AL560912
5	777	38.7	1061	10	BM464844	BM464844 AGENCOCURT
6	685.8	34.2	775	9	AL582241	AL582241 AL582241
7	659.8	32.9	769	10	BF038722	BF038722 601460641
8	654.2	32.6	702	10	BF038722	BF038722 602737137
9	651.4	32.5	888	10	BE869186	BE869186 601445180
10	651	32.4	784	10	BG120736	BG120736 602347389
11	650.2	32.4	752	10	BG682548	BG682548 602624362
12	632.6	31.5	908	10	BG420765	BG420765 602448558
13	632.4	31.5	668	10	B1758046	B1758046 603030577
14	629.8	31.4	645	9	AL559248	AL559248 AL559248
15	628	31.3	825	10	BE346959	BE346959 601071929
16	626.8	31.2	983	10	BE889971	BE889971 601512230
17	608.6	30.3	632	10	BF219707	BF219707 601296627

18	599.4	29.9	912	10	BG117979	BG117979 602351223
19	560.8	27.9	932	9	AL529568	AL529568 AL529568
20	556	27.7	619	9	AL580903	AL580903 AL580903
21	520.4	25.9	595	9	AL582265	AL582265 AL582265
22	498	24.8	702	10	B1151643	B1151643 602916189
23	490.4	24.4	905	9	AL563489	AL563489 AL563489
24	489.6	24.4	744	10	BF162255	BF162255 601770948
25	476.8	23.8	687	10	BF658170	BF658170 maag6b06.
26	462.8	23.1	523	9	AA564143	AA564143 h444f02.s
27	461.2	23.0	547	9	BE031265	BE031265 129770 MA
28	443	22.1	571	10	B1341090	B1341090 368355 MA
29	422.6	21.1	495	10	BG235999	BG235999 na2f0f09.
30	410.4	20.4	478	9	AM248701	AM248701 2821007.3
31	406.2	20.2	519	10	BE304189	BE304189 601086421
32	398.6	19.9	554	9	AA684194	AA684194 vm6810.s
33	387.4	19.3	589	10	BE334637	BE334637 us81g12.y
34	373.2	18.6	435	9	A1147333	A1147333 q963q01.s
35	365	18.2	455	9	AM207517	AM207517 UI-H-B11-
36	351.2	17.5	399	9	A1695169	A1695169 wa47h01.x
37	346	17.2	457	9	BE234038	BE234038 140628 MA
38	345.4	17.2	1015	10	BM462511	BM462511 AGENCOCURT
39	334	16.6	379	9	A1633797	A1633797 tt28f05.x
40	334	16.6	379	9	AM002888	AM002888 wt02c11.x
41	331.8	16.5	380	9	AM006305	AM006305 wq69h04.x
42	331.8	16.5	380	10	BE552288	BE552288 hy05e12.x
43	328.4	16.4	394	9	AM207547	AM207547 UI-H-B11-
44	327.6	16.3	370	9	AM249133	AM249133 2821007.5
45	321	16.0	366	9	A1699987	A1699987 tt35f05.x

ALIGNMENTS

RESULT 1	AK012948	LOCUS	AK012948	2696 bp	mus musculus 10, 11 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2810054N23.TRAF-Interacting protein, full insert sequence.	linear	HTC 19-JAN-2002
DEFINITION	AK012948	LOCUS	AK012948	2696 bp	mus musculus 10, 11 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2810054N23.TRAF-Interacting protein, full insert sequence.	linear	HTC 19-JAN-2002
ACCESSION	AK012948	VERSION	AK012948.1	GI:12850018			
KEYWORDS	HTC; CAP trapper.	SOURCE	Mus musculus (strain:C57BL/6J) 10, 11 days embryo cDNA to mRNA, clone:1b:RIKEN full-length enriched mouse cDNA library				
ORGANISM	Mus musculus						
REFERENCE	1 (sites)	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.						
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes						
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)						
MEDLINE	20499374						
PUBMED	11042159						
REFERENCE	3 (sites)						
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuami,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsunoto,H., Sakauchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,T., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.						
TITLE	RIKEN integrated sequence analysis (RISA) system-384 format						
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)						

MEDLINE	PUBLISHED	REFERENCE	AUTHORS
20530913	11/07/861	4 (sites)	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	JOURNAL	REFERENCE	AUTHORS
Functional annotation of a full-length mouse cDNA collection	Nature 409, 685-690 (2001)	5 (bases 1 to 2696)	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arkawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bulic, C., Carinini, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Horii, F., Hume, D., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Katoh, H., Kawaji, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishii, K., Nomura, K., Numata, I., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Salto, H., Salto, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schirml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sobabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, Y., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE	JOURNAL	COMMENT	
Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group; RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsukuba-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)		Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.	
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAAGAGAACAATCCACAGAGCTCTTTTGTGGTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 7.5 and subtraction to Rot = 37.5. Second strand cDNA was prepared with the primer adapted of sequence [5' GAGAGAGAGTCCTCGAGTAATTATAATTAATCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.			
FEATURES	SOURCE		
location/Qualifiers	1..2696		
/organism="Mus musculus"			
/strain="C57BL/6J"			
/db_xref="MGD:MGI:1896093"			
/db_xref="taxon:10090"			
/clone="2810054AN23"			
/clone_lib="RIKEN full-length enriched mouse cDNA library"			
/dev_stage="10, 11 days embryo"			
112..1524			
/gene="Traip"			
112..1524			
/gene="Traip"			
/note="TRAF-interacting protein data source:MGD, source key:MGI:1096377, evidence:ISS putative"			
codon_start=1			
/protein_id="BAB28567.1"			
/db_xref="GI:12850019"			
/translation="MPRLALDTTSDFPHSDRDVAIHCGTHFLQCLIQMFETAPSR TCPCRLQVKGKTIINKLFDLAEEENVDAEFLKNELDSVKQLSQOKREKDSQA IIPDRILBERNATVESLONALKNAEMLSCTLKQMKPLEODDETKAREAKRLK CKMTKMOIELLOSOSEVEEMIRDMGVSGAVAEOLAVYCSLKKEYNKKERAKD GELADRTKLDDLVSSRSKLKTINTELDAQIGLSAOKDLDSAOETSLRKMTIIQG TLSLPATNTNVSVRLVRESRPVPMNRPLHQPFGCEIDLTNTTFDVNTPTQTSGS CHLCPKLTLEPARSPMONVLAKHKVASPSQSGLGQRVGDELDAAGPFLFKNN AVLGQKPKNRTAYSRCSSTDPVRIGFGGLGRFTFIQPRDTIIRPVVSKASKSKOK			

[illegible]

OY	898	AGCTGMAAAGAAAGCTAAACGATGCTGCGAGGAAACCTTGAACCTTCCACCAAGTGGCCAGT	957
Db	907	AGCCCTAAGAAAGAAACCTGATGATCCTCCAGGGAAACCTTGAGGCTCTCCTCGGCCGACCAAT	966
OY	958	GAGATGTGTCGACGCGCTGGTTTTAGAGAGCCAGCCCGCTGTGCA---GGTGAATGTGAAG	1014
Db	967	GAGACGGTCAGCGCCCTGGTTTTTATAGAGCCAGCCCGCTGTGAGATGATGACCCGAGG	1026
OY	1015	CTCCGCGCGGCAATCCCTCCGTGATGATATATGATATCTCAATGCTACCTTTGATGTGATACT	1074
Db	1027	CTTACACAGCCACCTCGGTGATGAGATGTGATCTCAATACCACTTGATGATTAATACC	1086
OY	1075	CCCCAGCGCCGCTCCAGCTCCAGAGATGGTACTACGAAAACTTTGCTTAAGAAAG	1134
Db	1087	CTTCAACCCCAACCTCTGGCTCCAGAGATGGCTCCCCCAAAACCTGTGCTGGAGAGG	1146
OY	1135	TCACATCCCCCAATTCAGAGATGTCCCAAGAAAGATATGCAAAAGGGCCCCAGAGAGATCC	1194
Db	1147	GCACGCTCTCCCATGACAGATATCCTCAAGAAAGGTGCACAAAGTGTCCAGCCGGAGTCC	1206
OY	1195	CAGCTCTCACTGGGTGGCCAGAGCTGTGCGAGAGAGCCAGATGAGAGAACTGGTTGTGCC	1254
Db	1207	CAGCTCTCACTGGGTGGCCAGAGATGTGAGAGAGCTAGATGAGAGAACTGGCTGTGTGC	1266
OY	1255	TTCCCTATTTTTGTCCGGAATGCCATCCTAGGCCGAAAGACGCCCAAAAGCCCGCAGTCA	1314
Db	1267	TTCCCTCTCTTCATCCGGAAATGCTCTGCGGTGCGAAACACCCCAACAGGACCCACAGA	1326
OY	1315	GAGTCTCTGTGAGCAAAAGATGGTGAAGGACAGGCTTGATGGGCTGGTGGCCGGACA	1374
Db	1327	GAATCCCGTTGACGACAGATGTGTAGAAATAGGCTTTGATGGGCTGTGAGAGAGAACCA	1386
OY	1375	AAATTCATCCAGCTCACTAGTACACAGTCAATGATCCGCCCAATTCCTGTTAAAGCCCAAGAC	1434
Db	1387	AAATTCATCCAGCTCAAGGACACAAACACATATATCCGACAGTGCCTGTTAAAGTCCAAAGGC	1446
OY	1435	AAGGTTAAGCAGAGAGGTGAGGGTGAAGACCGTGCCCTTCTCTTCCAGGCCAAGCTGGAC	1494
Db	1447	AAGAGTAAACAAAGATGAGTAAGAACTGTGAGATCTGACCTCCAGGCCCAAGCTGGAT	1506
OY	1495	ACCTTCCTGTGTGCGAAGAACAGTAGAGTCTGACCAATGGCCAGACACATGCTCGCACT	1554
Db	1507	ACCTTCTTATGTGAC-----TGAAGGGTACCAAGGTCAATGTTTTGCAATT	1551
OY	1555	TGTAGGTCAAGGACTGTCCAGGAGGGTTTTGAGACAGAGCCCTACTTTCCGAGCAGAGCC	1614
Db	1552	ACTGGGCCCCAAGCTGGGTAAACCGG-----AAGTG	1581
OY	1615	TGAGGTGTAGGGCAGACAAACAGGTGAGGGTGAAGTGTGACACCCAGAGACTGCTCTTCC	1674
Db	1582	TTTTTTGGAAGATGGCTCCTTGGACACATCTCCAGAGAGATCCCAAGAAACACACTTCC	1641
OY	1675	TGCGCTCAACCTGCGCCCACTCTCAGACACTGGGAGGTGACATGACAGCCCACTGATCTCG	1734
Db	1642	TGTGTGCTACGTGGCCCTGTGAC--ACACTGGGAAGCCACATACCAAGTTTACTGTTCCGA	1699
OY	1735	TCAGAGAGTCTGC--TCTGTGTCACAGGCTTGTGTTATAGCCATGATCAGATGTGTGAG	1793
Db	1700	TCGGCAGGAGCTACTTCCAGGTTGCAAGGGTTTTGCTTATAGTACAAACAGGATGTGGCTGG	1759
OY	1794	ACTCTTTCTGGGCGCTGGAGACAGGCTCACTT	1825
Db	1760	ACTCCCTTTGTTTTATAGAACAGGAGTCAAT	1791

KEYWORDS HHC, CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) 10, 11 days embryo cDNA to mRNA, clone:2810021m06.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci, P. and Hayashizaki, Y.

JOURNAL High-efficiency full-length cDNA cloning

MEDLINE Meth. Enzymol. 303, 19-44 (1999)

PUBMED 99279253

REFERENCE 10349636

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20493574

PUBMED 11042159

REFERENCE 3 (sites)

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitanaka, T., Tashiro, H., Itoh, M., Suni, N., Ishii, T., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format

JOURNAL Sequencing pipeline with 384 multicapillary sequencer

MEDLINE Genome Res. 10 (11), 1757-1771 (2000)

PUBMED 11076861

REFERENCE 4 (sites)

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

MEDLINE 5 (bases 1 to 1938)

PUBMED 11076861

REFERENCE Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Butt, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Quekembush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schirral, L., Shibata, K., Shibata, Y., Shingagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for

RESULT	2
AK012786	
LOCUS	AK012786
DEFINITION	1938 bp mRNA linear HTC 19-JAN-2002
	Mus musculus 10, 11 days embryo whole body cDNA, RIKEN full-length
	enriched library, clone:2810021M06:TRAF-interacting protein, full
	insert sequence.

ACCESSION	AK012786
VERSION	AK012786.1
	GI:12849758

Encyclopedia of Life Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGAGATCCAGACCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 7.5 and subtraction to

and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 7.5 and subtraction to

Df		1560	-- -- -- -- --	-AAGTCTTTTGGAAGATGGCCTCTTGTGACCA	1592
OY	1643	GCGTAGTGTCACACCAGACTGCCTCTCCCTGCCCTCACCCCTCCCACATCTACAGC	1702		
Df	1593	GTCACAAGAGATGCCCCAGAATAAACACACTCTCTGTGTTCACATGCCCTGCACAC--ACAC	1650		
OY	1703	TGGGAGCTGACATTACACACAGCCCATCTGATCTCTTCACAGAGCTCTCAG-TCGTGTCCAGGC	1761		
Df	1651	TGGGAGGCCACATGACAGCTTACCTTCCTCCGATCTACAGAGGGCGCTCACTTCCACTCCAGGC	1710		
OY	1762	TCCTGTTTATTAACCCATGATCAGATGAGTCAGACATCTTCTGTGGCTGAGACACAGGC	1821		
Df	1711	TTTTCTTATTAACCTTACACACAGAGTGTGCTGAGCTCTCTTTCTTTTATTAAGACAGGGTC	1770		
OY	1822	AACTT 1825 Db 1771 ACAT 1774			
RESULT	3				
AL560947					
LOCUS	AL560947 LTI_NFL010_BC2 Homo sapiens cDNA clone CSODL005YM09 5	959 bp mRNA linear EST 16-FEB-2001			
DEFINITION	prime, mRNA sequence.				
ACCESSION	AL560947 GI:12907896				
VERSION	AL560947.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
I.(bases 1 to 959)					
L.I.W.B., Gruber,C., Jesssee,J. and Polayes,D.					
Full-length cDNA libraries and normalization unpublished (2001)					
JOURNAL COMMENT:	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr. Location/Qualifiers				
FEATURES	Source	1..959			
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone="CSODL005YM09"					
/clone_lib="LTI_NFL010_BC2"					
/sex="male"					
/tissue_type="B cells from Burkitt lymphoma"					
/note="Vector: pCMVSPORT 6; Site_: NotI; 1st strand CDNA was primed with a NotI-oligo(dt) primer. Five prime end enriched, double-stranded CDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@life-tech.com URL : http://fulllength.invitrogen.com"					
BASE COUNT	259 a 237 c 275 g 185 t	3 others			
ORIGIN					
Query Match	44.6%; Score 894.4; DB 9; Length 959; Best Local Similarity 97.6%; Pred. No. 5.le-157; Matches 937; Conservative 3; Mismatches 14; Indels 6; Gaps 3				
OY	20 GAAGCAAAGCGGAGCGGGCG-----CTTACGAAGCCGAGCTGTAGCAGTTCTTTGGC	75			
Df	1 GAAGCAAAGCGGAGCGGGCGGCGCTTACGAAGCGGAGCCGTGTAGCAGTTCTTTGGC	60			
OY	76 TGCCTGGGCCCCCTTGAGTCCAGCAATCATGCCATACCGTCCGTGACACTATTCGCTCC	135			
Df	61 TGCTGGGCCCCCTTGAGTCCAGCAATCATGCCATATCCGCTCTGTGCACTATCTCTCC	120			

QY	136	GACGTCCTCCGATCCACTCCCGGAGCTGGCGCCGATCCACTGGGCGCANACCTTCCACTTG	195
Db	121	GACCTCTTCGATCACTCCCGAGCGTGGCCGCATCCACTCGCGGCACACCTTCCACTTG	180
QY	196	CAGTGCCTAAATTCAGTCCCTTTGAGACACACCAAGTGCAGCTGCCACAGTGGCGAATC	255
Db	181	CAGTGCCTAAATTCAGTGGTTTGAGACACACCAAGTGCAGCTGCCACAGTGGCGAATC	240
QY	256	CAGGTTGGCAAAAGAACCATTAATCAATAGCTCTTTGATCTTGGCCAGAGAGAG	315
Db	241	CAGGTTGGCAAAAGAACCATTAATCAATAGCTCTTTGATCTTGGCCAGAGAGAG	300
QY	316	AATGCTCTGGATCCGAAATTCCTTAAAGATGAACGTGAGCATATGTCAGAGCCAGTTTCC	375
Db	301	AATGCTCTGGATCCGAAATTCCTTAAAGATGAACGTGAGCATATGTCAGAGCCAGTTTCC	360
QY	376	CAGAAAGACAAGAGAAACGAGACAGCCAGGCTCATCATCGACACTCTGCGGGATACGCTG	435
Db	361	CAGAAAGACAAGAGAAACGAGACAGCCAGGCTCATCATCGACACTCTGCGGGATACGCTG	420
QY	436	GAGAGAGCCATGCTACTGTGTATCTTGACAGCAGGCTTTGGGCABGGCCGAGATCTG	495
Db	421	GAGAGAGCCATGCTACTGTGTATCTTGACAGCAGGCTTTGGGCABGGCCGAGATCTG	480
QY	496	TGCTTCACACAGGAAAGAAACAGATGAAAGTACTTAGAGCAGCAGAGATGAGACCCAAACAA	555
Db	481	TGCTTCACACACTGAAAGAAACAGATGAAAGTACTTAGAGCAGCAGAGATGAGACCCAAACAA	540
QY	556	GCACAGAGAGAGGCGGCGCGCTCAGAGACAAGATGAAGAACCATGAGACAGATTGAGCTT	615
Db	541	GCACAGAGAGAGGCGCGCGCTCAGAGACAAGATGAAGAACCATGAGACAGATTGAGCTT	600
QY	616	CTACTCCAGAGCCAGCTCCCTGAGGTGGAGAGATGATCCGAGACATGGTGTGGGACAG	675
Db	601	CTACTCCAGAGCCAGCTCCCTGAGGTGGAGAGATGATCCGAGACATGGTGTGGGACAG	660
QY	676	TCAGGCGTGGAAACACACTGGCTGTGACTGTGTGCTCTCAAGAAAGAGTACGAGATCTA	735
Db	661	TCAGGCGTGGAAACACACTGGCTGTGACTGTGTGCTCTCAAGAAAGAGTACGAGATCTA	720
QY	736	AAAGAGGACACGGAGAGGCTCAGGGAGGTGGCTGACAAGCTGAGAGAGGATTTGTTTCC	795
Db	721	AAAGAGGACACGGAGAGGCTCAGGGAGGTGGCTGAC-ARCTGAGGAGAGGATTTGTTTCC	779
QY	796	TCCAAAGCAAGTTGCCACACAGTCTACTCTGATTTGGATCAGGCCCAATTTAGAACTGAG	855
Db	780	TCCAAAGCAAGTTGCCACACAGTCTACTCTGATTTGGATCAGGCCCAATTTAGAACTGAG	839
QY	856	TCAGCCCAAGAGGATTTACAGAGTCTGTACAAGAAATCATGAGGCTGAAAAAGAACCTA	915
Db	840	TCAGCCCAAGAGGATTTACAGAGTCTGTACAAGAAATCATGAGGCTGAAAAAGAACCTA	899
QY	916	ACGATGCTGCAGAAACCTT-GAACCTGCCACACAGTGGCCAGTGGAGACTGTGCAGCGCTT	974
Db	900	ACGATGCTGCAGAAACCTTGGAAACCTGCCACACAGTGGGSCAGTGGAGACTGTGCAGCGCTT	959
RESULT 4			
AL560912		887 bp	mRNA
LOCUS		AL560912	887 bp
DEFINITION		AL560912 LTI-NFL010-BC2 Homo sapiens cdna clone CS0DL005YI08	5
ACCESSION		AL560912	prime, mRNA sequence.
VERSION		AL560912.1	GI:12907828
KEYWORDS		EST.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		Li,W.B., Gruber,C., Jeesee,J. and Polayes,D.	
		1 (bases 1 to 887)	
		Full-length cDNA libraries and normalization	

JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.

FEATURES
 source
 1. .887
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DL005Y108"
 /clone_lib="LTI-NFL010_BC2"
 /sex="male"
 /tissue_type="B cells from Burkitt lymphoma"
 /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com ;

BASE COUNT 251 a 219 c 245 g 171 t 1 others

ORIGIN

Query Match 43.0%; Score 863.4; DB 9; Length 887;
 Best Local Similarity 99.0%; Pred. No. 3.1e-151;
 Matches 878; Conservative 1; Mismatches 7; Indels 1; Gaps 1;

84 CCCCTTGAGTCCAGCATCATGCTATCCGCTCTGTGCACTATGCTCCGACTTCTT 143
 Db 1 CCCCTTGAGTCCAGCATCATGCTATCCGCTCTGTGCACTATGCTCCGACTTCTT 60

144 CGATCACTCCCGGACGTGGCCGCACTCGCGGCCACACTTCCACTTGGAGTGGCT 203
 Db 61 CGATCACTCCCGGACGTGGCCGCACTCGCGGCCACACTTCCACTTGGAGTGGCT 120

204 AATTCAGTCTTTGAGACAGCAGCAAGTCGAGCTGGCCACAGTGGCAATCCAGGTTGG 263
 Db 121 AATTCAGTCTTTGAGACAGCAGCAAGTCGAGCTGGCCACAGTGGCAATCCAGGTTGG 180

264 CAAAAGAACCATTTATCAATAAGCTCTTCTTGATCTTGGCCAGAGAGAGAAATGTTCTT 323
 Db 181 CAAAAGAACCATTTATCAATAAGCTCTTCTTGATCTTGGCCAGAGAGAGAAATGTTCTT 240

324 GGATCGAATTTCTTAAAGATGAACTGGACAATGTCTAGAGCCCAAGCTTTCCAGAAAGA 383
 Db 241 GGATCGAATTTCTTAAAGATGAACTGGACAATGTCTAGAGCCCAAGCTTTCCAGAAAGA 300

384 CAAGGAGAAACGAGACAGCGCATCATGCACTCTGGGGATGACGGTGGAAAGC 443
 Db 301 CAAGGAGAAACGAGACAGCGCATCATGCACTCTGGGGATGACGGTGGAAAGC 360

444 CAATGCTACTGTGTATCTCTGACAGAGCCCTTGGGCAAGGCCAGATGCTGTCTCCAC 503
 Db 361 CAATGCTACTGTGTATCTCTGACAGAGCCCTTGGGCAAGGCCAGATGCTGTCTCCAC 420

504 ACTAAAAAGCAGATGAAGTACTTGAACAGCAGCAGAGATGAGACCAACAGCAGACAGA 563
 Db 421 ACTAAAAAGCAGATGAAGTACTTGAACAGCAGCAGAGATGAGACCAACAGCAGACAGA 480

564 GGAGGCGGGCGGCTCAGAGACAGATGAAGACCATGAGAGATGAGCTTCACTCCA 623
 Db 481 GGAGGCGGGCGGCTCAGAGACAGATGAAGACCATGAGAGATGAGCTTCACTCCA 540

624 GAGCAGACTCCCTGAGTGGAGAGATGATCCAGACATGGGTGTGGAGACAGTACGGGT 683
 Db 541 GAGCAGAGGCGCTGAGTGGAGAGATGATCCAGACATGGGTGTGGAGACAGTACGGGT 600

684 GGAACAGCTGGCTGTACTGTGTCTTCACAGAAAGATGACGAATCTTAAAGAGGC 743
 |||||||

Db 601 GGAACAGCTGGCTGTACTGTGTCTCTCAAGAAAGATAGCAGAAATCTAAAGAGGC 660

744 ACGAAGGCCCTCAGAGGAGGTGGCTGACAGCTG -ACGAGAGATTGTCTCCCGAA 802
 |||||||

661 ACGAAGGCCCTCAGAGGAGGTGGCTGACAGCTGAAGCTGAAGAGATTGTCTCCCGAA 720
 |||||||

803 GCAAGTTCAGACAGTCTACTCTGAATTTGATGATGAGGCCAAGTTAGACTAGAGCC 862
 Db 721 GCAAGTTCAGACAGTCTACTCTGAATTTGATGATGAGGCCAAGTTAGACTAGAGCC 780

863 AGAAGCACTTACAGAGTCTGACAGAGAAATCATGAGCCTGAAAAAGACTTAACATGC 922
 Db 781 AGAAGCACTTACAGAGTCTGACAGAGAAATCATGAGCCTGAAAAAGACTTAACATGC 840

923 TGCAGGAACCTTGAACCTGCCACACAGTGGCCAGTGCAGTGCAGC 969
 |||||||

841 TGCAGGAACCTTGAACCTGCCACACAGTGGCCAGTGCAGTGCAGC 887

RESULT 5
 LOCUS BM464844 1061 bp mRNA EST 05-FEB-2002
 DEFINITION AGENCOURT_6428661 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5504946
 5', mRNA sequence.
 ACCESSION BM464844
 VERSION BM464844.1 GI:18513886
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 1061)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabbs@email.nih.gov
 Tissue Procurement: Lou Staudt
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12147 row: b column: 19
 High quality sequence start: 2
 High quality sequence stop: 641.

FEATURES
 source
 1. .1061
 Location/Qualifiers
 1. .1061
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5504946"
 /clone_lib="NIH_MGC_85"
 /tissue_type="lymphoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by life technologies. Note: this is a NIH_MGC Library."

BASE COUNT 282 a 272 c 297 g 202 t 8 others

ORIGIN

Query Match 38.7%; Score 777; DB 10; Length 1061;
 Best Local Similarity 97.2%; Pred. No. 3.9e-135;
 Matches 808; Conservative 0; Mismatches 21; Indels 2; Gaps 2;

34 CGGGGGCTCTAGAGCGGACGCTAGACATTTCTTTGGCTGGCCCTGGAGT 93
 |||||||

26 CGGGCGGCTTACAGAGCGGACCTGTAGAGATTTCTTTGGCTGGCCCTGGAGT 85
 |||||||

94 CCAGCATCATGCTATCCGCTGTGTGACATATCTGCTCCGACTTCTTGATCACTCC 153
 |||||||

```

Db      86 CCACCCATCATGCTATCCGCTCTGCGACATCTGCTCCGACTTCTTCGATCACTCC 145
QY      154 CGGACAGCGGCGGCATCCACTGCGGCACACCTTCACATTGCGATGCTAATTCAGTCC 213
Db      146 CGGACAGCGGCGGCATCCACTGCGGCACACCTTCACATTGCGATGCTAATTCAGTCC 205
QY      214 TTGTAGACAGCACCAGTCCGACCTGCGCCACAGTCCGCAATCCAGTTGGCAAAAGAAC 273
Db      206 TTGTAGACAGCACCAGTCCGACCTGCGCCACAGTCCGCAATCCAGTTGGCAAAAGAAC 265
QY      274 ATTATCAATAAGCTCTTCTTTGATCTTGGCCAGAGAGAGAAATGCTTGGATCGAAA 333
Db      266 ATTATCAATAAGCTCTTCTTTGATCTTGGCCAGAGAGAGAAATGCTTGGATCGAAA 325
QY      334 TTCTTAAGATGATGACGAGCAATGTCAGAGCCAGCTTCCCAAGAAAGCAAGAGAAA 393
Db      326 TTCTTAAGATGATGACGAGCAATGTCAGAGCCAGCTTCCCAAGAAAGCAAGAGAAA 385
QY      394 CGAGACAGCCAGTCAATCATGACACTCTGCGGGATACGCTGGAAGAAGCAATGCTACT 453
Db      386 CGAATACAGCCAGTCAATCATGACACTCTGCGGGATACGCTGGAAGAAGCAATGCTACT 445
QY      454 GTGATATCTCTGACAGAGGCTTGGGCAAGGCCGAGATGCTGCTCCACACTGAAAAAG 513
Db      446 GTGATATCTCTGACAGAGGCTTGGGCAAGGCCGAGATGCTGCTCCACACTGAAAAAG 505
QY      514 CAGATGAAGTACTTAGAGCAGCAGAGATGAGACCAACAGCAAGAGAGAGCGGGC 573
Db      506 CAGATGAAGTACTTAGAGCAGCAGAGATGAGACCAACAGCAAGAGAGAGCGGGC 565
QY      574 CGGCTCAGAGCAGCAGATGAAAGACCATGAGAGATGAGCTTCTTACATCCAGCAGCAGTC 633
Db      566 CGGCTCAGAGCAGCAGATGAAAGACCATGAGAGATGAGCTTCTTACATCCAGCAGCAGTC 625
QY      634 CCTGAGGTGAGAGAGATGATCCGAGACATGAGTGTGGACACATCGCGGTGGAACAGTGT 693
Db      626 CCTGAGGTGAGAGAGATGATCCGAGACATGAGTGTGGACACATCGCGGTGGAACAGTGT 685
QY      694 GCTGTGATCTGTGTCTCTCTAAGAAGAGTACGAGAAATCTAAAAAGAGCGCAAGAGCC 753
Db      686 GCTGTGATCTGTGTCTCTCTAAGAAGAGTACGAGAAATCTAAAAAGAGCGCAAGAGCC 745
QY      754 TCAGGGAGAGTGGCTGACAAAGCTGAGAAAGATTTGTTTCC-TCAGAAACCAAGTTGCA 812
Db      746 TCAGGGAGAGTGGCTGACAAAGCTGAGAAAGATTTGTTTCC-TCAGAAACCAAGTTGCA 805
QY      813 GACAGTCTACTCTGAATTGGATCAGGCCAAG-TTAGAACTGAGTCAAGCCC 862
Db      806 GACAGTCTACTCTGAATTGGATCAGGCCAAGTTTAGAACTGAGTCAAGCCC 856

RESULT 6
AL582241/c 775 bp mRNA linear EST 16-FEB-2001
LOCUS      AL582241 LTL_NFL010_BC2 Homo sapiens cDNA clone CS0DL005Y108 3
DEFINITION prime, mRNA sequence.
ACCESSION AL582241
VERSION    AL582241.1 GI:12950030
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 775)
            Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
            Full-length cDNA libraries and normalization
            JOURNAL Unpublished (2001)
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.
FEATURES
            Location/Qualifiers

```

```

source
1. .775
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DL005Y108"
/clone_id="LTL_NFL010_BC2"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/issue="Vector: PCWVSORT 6; Site: 1; Note: 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCWVSORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Jiang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fjiang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 156 a 199 c 217 g 192 t 11 others
ORIGIN

Query Match 34.2%; Score 685.8; DB 9; Length 775;
Best Local Similarity 95.2%; Pred. No. 4.1e-118;
Matches 738; Conservative 11; Mismatches 18; Indels 8; Gaps 4;

QY 1045 GATCTCAATGCTACTCTTGAATGATACGCTCCCGCCGCGCCCTCCAGCTCCAGCAT 1104
Db 775 GATCTCAATGCTACTCTTGAATGATACGCTCCCGCCGCGCCCTCCAGCTCCAGCAT 716
QY 1105 GGTCTACTACGAAAAAATTTGGCTGAGAGAGTCACTCTCCCAATTCAGATGTGCCCAAG 1164
Db 715 GGTCTACTACGAAAAAATTTGGCTGAGAGAGTCACTCTCCCAATTCAGATGTGCCCAAG 656
QY 1165 AAGATATGCAAAAGCCCGAGAGAGAGTCCAGCTCTCACTGGTGGCCAGAGCTGTGCA 1224
Db 655 AAGATATGCAAAAGCCCGAGAGAGAGTCCAGCTCTCACTGGTGGCCAGAGCTGTGCA 596
QY 1225 GGAAGGCGAGATGAGAAAGTGTGGTGGCTTCCCTTTTGGTCGGAAATGCCATCTTA 1284
Db 595 GGAAGGCGAGATGAGAAAGTGTGGTGGCTTCCCTTTTGGTCGGAAATGCCATCTTA 536
QY 1285 GGCAGGACCAAGTGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1344
Db 535 GGCAGGACCAAGTGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 476
QY 1345 ACAGGCTTCATGAGGCTCGGTTGGCCGAGCAAAATTCATCCAGCTTACTGACACAGTATG 1404
Db 475 ACAGGCTTCATGAGGCTCGGTTGGCCGAGCAAAATTCATCCAGCTTACTGACACAGTATG 416
QY 1405 ATCCGCCCATTTGCTTTAAGCCCAAGCAAGGTTAAGCAGAGAGGTGAGAGTGAAGCC 1464
Db 415 ATCCGCCCATTTGCTTTAAGCCCAAGCAAGGTTAAGCAGAGAGGTGAGAGTGAAGCC 356
QY 1465 GTGCTTCTCTCTTCCAGGCGCAAGCTGAGACCTTCTGTGGTGTGAGAGACAGTGAATC 1524
Db 355 GTGCTTCTCTCTTCCAGGCGCAAGCTGAGACCTTCTGTGGTGTGAGAGACAGTGAATC 296
QY 1525 TGACCAATGGCCCAACACATGCTGCAACTTTGATGTTCAAGAGATGTGACAGGCG--T 1582
Db 295 TGACCAATGGCCCAACACATGCTGCAACTTTGATGTTCAAGAGATGTGACAGGCGGTT 236
QY 1583 TTGTGACAGAGCCCTTCTTGGGACAGCCTGAGGTGAAGGGCAGACAAACAGGTGA 1642
Db 235 TTGTGACAGAGCCCTTCTTGGGACAGCCTGAGGTGAAGGGCAGACAAACAGGTGA 176
QY 1643 GGGTGAAGTGAACACCCAGAGACAGTCTTCTCCGCTCCACCTCCACCTCCCTAGAC 1702
Db 175 GGGTGAAGTGAACACCCAGAGACAGTCTTCTCCGCTCCACCTCCACCTCCCTAGAC 116
QY 1703 TGGAGCTGACAT-GACCAGCCCACTGATCTGTACAGAGTCTCTGCTC---TGTGGCC 1757
Db 115 TGGAACTGAAATAGACACACCCAMMGATCTCTGTACAGAGTCTCTGCTGATAGCACA 56

```


QY	1758	AGGCTCTGTTTATAGC-CATGATAGATGTGTGTAAGACTCTTTCTGGCCCTGGA	1811
Db	55	GGACTTCGTTTATAGCACAGATGATGTGTGTAAGACTCTTTCTGGCCCTRAA	1
RESULT	7		
LOCUS	BF038722	769 bp	mRNA
DEFINITION	601460641P1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3864082 5',		linear EST 20-OCT-2000
ACCESSION	BF038722		mRNA sequence.
VERSION	BF038722.1	GI:10745770	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NIH-MGC htcp://mhc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strusberg, Ph.D.		
	Email: c9apbs-rf@mail.nih.gov		
	Tissue Procurement: DCTD/DP		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		
	Plate: LLM9605 row: a column: 11		
	High quality sequence stop: 632.		
FEATURES	Location/Qualifiers		
source	1..769		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3864082"		
	/clone_lid="NIH_MGC_66"		
	/tissue_type="adenocarcinoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;		
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.		
	Average insert size 1.8 kb. Library constructed by Life		
	Technologies."		
BASE COUNT	184 a 212 c 217 g 156 t		
ORIGIN			
Query Match	32.9%; Score 659.8; DB 10; Length 769;		
Best Local Similarity	97.8%; Pred. No. 2.9e-113;		
Matches 701; Conservative	0; Mismatches 12; Indels 4; Gaps 3;		
QY	930	AACCTTGAAACCGCCACCAAGTCCGACAGTGAAGTGTGACCCCGCTGTTTATAGAGAGCC	989
Db	1	AACCTTGAAACCGCCACCAAGTCCGACAGTGAAGTGTGACCCCGCTGTTTATAGAGAGCC	60
QY	990	AGCCCCCTGGAGAGTGAATCTGAAGCTCCGCGGCGCATCTTCGATGATATGATCT	1049
Db	61	AGCCCCCTGGAGAGTGAATCTGAAGCTCCGCGGCGCATCTTCGATGATATGATCT	120
QY	1050	CAATGCTACCTTTGATGTGATAGTACCTCCCAAGCCCGGCCCTCCAGCTCCAGCATGTTA	1109
Db	121	CAATGCTACCTTTGATGTGATAGTACCTCCCAAGCCCGGCCCTCCAGCTCCAGCATGTTA	180
QY	1110	CTACGAAAAAATCTTGCTCTGAGAGAATGTACACCTCCCAATTCAGATGTGCCAAGAAGT	1169
Db	181	CTACGAAAAAATCTTGCTCTGAGAGAATGTACACCTCCCAATTCAGATGTGCCAAGAAGT	240
QY	1170	ATGCAAAAGGCCCCAGGAAGAGTCCAGCTCAGTGGGTGGCCAGAGCTGTGCAGGAGA	1229
Db	241	ATGCAAAAGGCCCCAGGAAGAGTCCAGCTCAGTGGGTGGCCAGAGCTGTGCAGGAGA	300
QY	1230	GCCAGATGAGAACTGTTGTGCTTCCTCTTTTGTGTCCGAATGCCATCTGAGCCA	1289

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT					
Db	301	GCACATGAGGAACGTGGTGGCTTCCCTTAATTTTGTCCGGAATGCCATCTAGGCGCA	1290	GAACAGGCCAAAAGAGGCCAGTCAGAGTCTCTTGCAGCAAAAGATGTGTAAAGACAGG	1349	Db	361	GAACAGGCCAAAAGAGGCCAGTCAGAGTCTCTTGCAGCAAAAGATGTGTAAAGACAGG	420	Qy	1350	CTTCGATGGGCTCGGTGGCCGGACAAAATTCATCCAGCTTACTGACACAGTCATGATCCG	1409				
Db	421	CTTCGATGGGCTCGGTGGCCGGACAAAATTCATCCAGCTTACTGACACAGTCATGATCCG	1410	CCCATTTGGCTGTAAAGCCCAAGACCAAGTTAAGCAGAGGGGTGAGGGTGAAGACCTGTGCC	1469	Db	481	CCCATTTGGCTGTAAAGCCCAAGACCAAGTTAAGCAGAGGGGTGAGGGTGAAGACAGTGTCC	540	Qy	1470	TTTCTCTTTCAGAGCCAGTCAGACACCTTCCTGTGGTGTGAGAACAGTGAAGTGTGACC	1529				
Db	541	TTTCTCTTTCAGAGCCAGTCAGACACCTTCCTGTGGTGTGAGAACAGTGAAGTGTGACC	541	TTTCTCTTTCAGAGCCAGTCAGACACCTTCCTGTGGTGTGAGAACAGTGAAGTGTGACC	600	Db	601	AATGGCAGACACATGCTGTCAACTTGTAGGTCAAGAGACTGTCCAGGAGG--TTTGTG	1587	Qy	1530	AATGGCAGACACATGCTGTCAACTTGTAGGTCAAGAGACTGTCCAGGAGG--TTTGTG	1587				
Qy	1588	GACACAGCCCTTACTTTCGGGACACAGCTGTAGGTGTAAAGGACAGACAAACAGTGTAGG	1644	GACACAGCCCTTACTTTCGGGACACAGCTGTAGGTGTAAAGGACAGACAAACAGTGTAGG	1644	Db	660	GCCACAGGCCCAACTTTCGGGAAACAG-CTGAGGTGTGTAGGAGGACACACAGTGTGACG	715	RESULT	8	BG764118	702 bp	MRNA	Linear	EST 15-MAY-2001	
LOCUS	BG764118	602737137F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4862312 5'	BG764118	GI:14074771	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	1 (bases 1 to 702)	NIH-MGC http://mgc.nci.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)	Contact: Robert Strausberg, Ph.D. Email: rgs@biml.nih.gov Tissue Procurement: ATCC/DCTP/DRP	cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)	DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov	Plate: LUCM1722 row: 1 column: 09	High quality sequence stop: 697.	Location/Qualifiers
FEATURES	source	1..702	/organism="Homo sapiens"	/db_xref="taxon.9606"	/clone="IMAGE:4862312"	/clone_lib="NIH_MGC_49"	/tissue_type="melanotic melanoma, high MDR (cell line)"	/lab_host="DH10B (phage-resistant)"	/note="Organ: skin; Vector: pCMV7; Site: 1; XhoI: Site 2; EcoRI: cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCGCAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH_MGC Library."	183 a	180 c	202 g	136 t	1 others			

Query Match	32.4%	Score 651;	DB 10;	Length 784;
Best Local Similarity	93.6%	Pred. No. 1,3e-111;		
Matches 735; Conservative	0;	Mismatches 40;	Indels 10;	Gaps 5;
QY 16	ATTTCAGACGAGCGGAGCC-----GGGGCGCTCTACGAAGCGGACCTGTAGCAATTTCCT			
Db 1	ATTTCAGACGAGCGGAGCGCGGGGCGCGTCTACGAAGCGCGGACCTGTAGCAATTTCCT			
QY 71	TTGGCTGCCTGGGCCCTTGAGTCCAGCCATCATGCCATATCCGTGCTGTGGCACTACTCT			
Db 61	TTGGCTGCCTGGGCCCTTGAGTCCAGCCATCATGCCATATCCGTGCTGTGGCACTACTCT			

RESULT 11
BG682548
LOCUS
DEFINITION 752 bp mRNA linear EST 01-MAY-2001
ACCESSION BG682548
VERSION BG682548
KEYWORDS BG682548.1 GI:13913945
SOURCE EST.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 752)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be


```
LOCUS      AL559248      645 bp      mRNA      linear      EST 16-FEB-2001
DEFINITION AL559248 LTI_NFL008.TC2 Homo sapiens cDNA clone CS0DJ012YH21 5
prime, mRNA sequence.
ACCESSION  AL559248
VERSION     AL559248.1  GI:12904563
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 645)
AUTHORS   Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: segrete@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES   source
            1. 645
                Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="CS0DJ012YH21"
                /clone_lib="LTI_NFL008_TC2"
                /sex="male"
                /tissue.type="T cells from T cell leukemia"
                /note="Vector: pCMVSPORT 6; Site.1: NotI; 1st strand cDNA
                was primed with a NotI-oligo(dT) primer. Five prime end
                enriched, double-stranded cDNA was digested with Not I and
                cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                vector. Library was normalized. Library was constructed by
                Life Technologies. Contact : Feng liang Life Technologies,
                a division of Invitrogen 9800 Medical Center Drive
                Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                Email : fliang@lifestech.com URL :
                http://fulllength.invitrogen.com"
BASE COUNT  167 a 174 c 177 g 127 t
ORIGIN
Query Match      31.4%; Score 629.8; DB 9; Length 645;
Best Local Similarity 98.9%; Pred. No. 1.2e-107;
Matches 634; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 42 TCTACGAGCGGACCGCTGACAGTTCTTGGCTGCGGCCCCCTTGATCCAGCCAT 101
DB 5 TCTACGAGCGGACCGCTGACAGTTCTTGGCTGCGGCCCCCTTGATCCAGCCAT 64
QY 102 CATGCCATATCGTCTGTGCACTATCTCGACTTCTTGATCATCTCCGCGAGCT 161
DB 65 CATGCCATATCGTCTGTGCACTATCTCGACTTCTTGATCATCTCCGCGAGCT 124
QY 162 GGGCGGCATCCACTGGGGCCACACCTTCCACTGGAGTGCCTAATTCAGTCTTTGAGAC 221
DB 125 GGGCGGCATCCACTGGGGCCACACCTTCCACTGGAGTGCCTAATTCAGTCTTTGAGAC 184
QY 222 AGCACCAAGTCGAGACCGCCACAGTCCGAATCCAGGTTGGCAAAAGACCATTAACA 281
DB 185 AGCACCAAGTCGAGACCGCCACAGTCCGAATCCAGGTTGGCAAAAGACCATTAACA 244
QY 282 TAAGCTCTTTTGTGATCTTCCAGAGAGAGAAATGTTGGATGAGAAATTTCTTAAA 341
DB 245 TAAGCTCTTTTGTGATCTTCCAGAGAGAGAAATGTTGGATGAGAAATTTCTTAAA 304
QY 342 GAATGAATGAGACAATGTCAAGAGCCAGCTTTCCAGAAAAGACAAGAGAAAGAGACAG 401
DB 305 GAATGAATGAGACAATGTCAAGAGCCAGCTTTCCAGAAAAGACAAGAGAAAGAGACAG 364
QY 402 CCAAGTCATCATGACACTGTGGGGATAGCTGGAAAGAAAGCAATCTCTGTTGATC 461
DB 365 CCAAGTCATCATGACACTGTGGGGATAGCTGGAAAGAAAGCAATCTCTGTTGATC 424
QY 462 TCTGCAGACAGCCCTGGGGCAAGCCGAGATCTGTCTCCACACTGAAAAAGAGATGAA 521
```

```
|||||
DB 425 TCTGCAGACAGCCCTGGGGCAAGGCCGAGTGTCTGTCTCCACTGAAAAAGAGATGAA 484
QY 522 GTACTTAGACAGCAGCAGATGAGACCAAAACAAGACAAGAGAGGCGCGGCTCAG 581
DB 485 GTACTTAGACAGCAGCAGATGAGACCAAAACAAGACAAGAGAGGCGCGGCTCAG 544
QY 582 GAGCAAGATGAAGACCTGGAGCAGATTGAGCTTCTCTCCAGAGCCAGCTCCCTGAGGT 641
DB 545 GAGCAAGATGAAGACCTGGAGCAGATTGAGCTTCTCTCCAGAGCCAGCTCCCTGAGGT 604
QY 642 GGAAGAGATGATCCGAGACATGGGTGTGGACAGTCACCG 682
DB 605 GGAAGAGATGATCCGAGACATGGGTGTGGACAGTCACCG 645
RESULT 15
BE546959
LOCUS      BE546959      825 bp      mRNA      linear      EST 09-AUG-2000
DEFINITION 601071929F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3458023 5',
mRNA sequence.
ACCESSION  BE546959
VERSION     BE546959.1  GI:9775604
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 825)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabos-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: L14M8448 row: j column: 08
            High quality sequence stop: 655.
FEATURES   source
            1. 825
                Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3458023"
                /clone_lib="NIH_MGC_12"
                /tissue.type="cervical carcinoma cell line"
                /lab_host="DH10B"
                /note="Organ: cervix; Vector: pCMV-SPORT6; Site.1: NotI;
                Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.4 kb. Library prepared by Life
                Technologies."
BASE COUNT  230 a 209 c 242 g 144 t
ORIGIN
Query Match      31.3%; Score 628; DB 10; Length 825;
Best Local Similarity 95.6%; Pred. No. 2.4e-107;
Matches 711; Conservative 0; Mismatches 25; Indels 8; Gaps 6;
QY 126 TATCTGCTCGACTTCTTGATCATCTCCGCGAGTGGCGCCATCCAGTGGCGCACAC 185
DB 1 TATCTGCTCGACTTCTTGATCATCTCCGCGAGTGGCGCCATCCAGTGGCGCACAC 60
QY 186 CTTCACCTTCAGAGCGCTTAATTCAGTCTTGAGACGACCAAGTGGAGCTGGCCACA 245
DB 61 CTTCACCTTCAGAGCGCTTAATTCAGTCTTGAGACGACCAAGTGGAGCTGGCCACA 120
QY 246 GTGCCAATCCAGTTGGCAAAAGAACATTATCAATTAAGCTCTTCTTATCTTGCCCA 305
|||||
```


mRNA sequence.
 ACCESSION BE219707
 VERSION BE219707.1 GI:11125801
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 632)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
 Plate: IRAL1 row: e column: 07
 High quality sequence stop: 630.
 Location/Qualifiers
 1..632
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2821007"
 /clone_1b="NIH-MGC-7"
 /tissue_type="small cell carcinoma"
 /cell_line="MGC3"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: POTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGCAG(C). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 163 a 171 c 172 g 126 t
 ORIGIN
 Query Match 30.3%; Score 608.6; DB 10; Length 632;
 Best Local Similarity 98.6%; Pred. No. 1.e-103;
 Matches 614; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 Oy 42 TCTACGAAGCGGACCTGTAGCACTTCTTGCTGCTGGCCCTTGAGTCCAGCCAT 101
 |||||||
 Db 7 TCTACGAAGCGGACCTGTAGCACTTCTTGCTGCTGGCCCTTGAGTCCAGCCAT 66
 Oy 102 CATGCCATCGGCTGTGACACTATCTGCTCGACTTCTTGATCATGCCGCGAGCT 161
 |||||||
 Db 67 CATGCCATCGGCTGTGACACTATCTGCTCGACTTCTTGATCATGCCGCGAGCT 126
 Oy 162 GGCGCCATCCACTGCGGCGACACACTTCCACTTGCTTAATTCAGTCTTTGAGAC 221
 |||||||
 Db 127 GGCGCCATCCACTGCGGCGACACACTTCCACTTGCTTAATTCAGTCTTTGAGAC 186
 Oy 222 AGACCAAGTGGACCTGCCACAGTCCGATCCAGGTTGGCAAAAGCAATATCA 281
 |||||||
 Db 187 AGCACCAAGTGGACCTGCCACAGTCCGATCCAGGTTGGCAAAAGCAATATCA 246
 Oy 282 TAAGCTCTTCTTGATCTTGCGCGAGGAGGAGATGCTTGATGAGATCTTAA 341
 |||||||
 Db 247 TAAGCTCTTCTTGATCTTGCGCGAGGAGGAGATGCTTGATGAGATCTTAA 306
 Oy 342 GAATGAATGACATGTACAGAGCCAGCTTTCCAGAAAGCAAGAGAGAGAGAG 401
 |||||||
 Db 307 GAATGAATGACATGTACAGAGCCAGCTTTCCAGAAAGCAAGAGAGAGAGAG 366
 Oy 402 CCAGGTATCATGACACTTGGGGGATACGCTGAAAGAACGCAATCTACTGTGATC 461
 |||||||
 Db 367 CCAGGTATCATGACACTTGGGGGATACGCTGAAAGAACGCAATCTACTGTGATC 426

Oy 462 TCTGACAGACGCTTGGGCAAGCCGAGATGCTGTGCTCCACACTGAAGAGATGAA 521
 |||||||
 Db 427 TCTGACAGACGCTTGGGCAAGCCGAGATGCTGTGCTCCACACTGAAGAGATGAA 486
 Oy 522 GTACTTAGACAGCAGGAGATGAGACCAACAGCAAGAGAGGCGGCGGCTCAG 581
 |||||||
 Db 487 GTACTTAGACAGCAGGAGATGAGACCAACAGCAAGAGAGGCGGCGGCTCAG 546
 Oy 582 GAGCAAGATGAAGACCATGAGAGATTTACTCTCAGAGCCAGCTCCCTAGGT 641
 |||||||
 Db 547 GAGCAAGATGAAGACCATGAGAGATTTACTCTCAGAGCCAGCTCCCTAGGT 606
 Oy 642 GGAGAGATGATCCGAGACATGG 664
 |||||||
 Db 607 GGAGAGATGATCCGAGACTGG 629
 RESULT 18
 BG117979 912 bp mRNA linear EST 30-JAN-2001
 LOCUS BG117979
 DEFINITION 602351223F1 NIH-MGC_90 Homo sapiens cDNA clone IMAGE:4446149 5',
 mRNA sequence.
 ACCESSION BG117979
 VERSION BG117979.1 GI:12611485
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 912)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
 Plate: LLAM10224 row: n column: 06
 High quality sequence stop: 678.
 Location/Qualifiers
 1..912
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4446149"
 /clone_1b="NIH-MGC-90"
 /tissue_type="adenoecarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies.
 Note: this is a NIH-MGC library."
 BASE COUNT 236 a 243 c 273 g 160 t
 ORIGIN
 Query Match 29.9%; Score 599.4; DB 10; Length 912;
 Best Local Similarity 91.3%; Pred. No. 5.2e-102;
 Matches 737; Conservative 0; Mismatches 56; Indels 14; Gaps 9;
 Oy 14 AAATTGAAGCAAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 68
 |||||||
 Db 2 AAATTGAAGCAAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61
 Oy 69 CTTTGCGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 128
 |||||||
 Db 62 CTTTGCGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 121

prime, mRNA sequence.
ACCESSION AL580903
VERSION AL580903.1 GI:12947374
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 619)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1. 619
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DJ012YH21"
/clone_11b="LTI_NFL008_Tc2"
/sex="male"
/issue_type="T cells from T cell leukemia"
/note="Vector: PCWVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCWVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 134 a 171 c 176 g 136 t 2 others
ORIGIN

Query Match 27.7%; Score 556; DB 9; Length 619;
Best Local Similarity 97.7%; Pred. No. 6.8e-94;
Matches 604; Conservative 2; Mismatches 7; Indels 5; Gaps 4;

QY 1285 GGGCAGAAACAGCCCAAAAGCCAGCTGAGTCCCTTGCAGCAAAAGATGTGTAAAG 1344
DB 619 GGGCAGAAACAGCCCAAAAGCCAGCTGAGTCCCTTGCAGCAAAAGATGTGTAAAG 560
QY 1345 ACAAGCTTCAGTGGCTGGTGGCCGACAAATTCATCCAGCCTACTGACACAGTCATG 1404
DB 559 ACAAGCTTCAGTGGCTGGTGGCCGACCAAAATTCATCCAGCCTACTGACACAGTCATG 500
QY 1405 ATCCGCCATTCGCTGTAAAGCCCAAGGTTAAGCAGAGGTTGAGGTTGAAGACC 1464
DB 499 ATCCGCCATTCGCTGTAAAGCCCAAGGTTAAGCAGAGGTTGAGGTTGAAGACA 440
QY 1465 GTGCTCTCTCTTCCAGGCAAGCTGACACCTTCCTGTGTCGTGAGAAACAGTGAAGTC 1524
DB 439 GTGCTCTCTCTTCCAGGCAAGCTGACACCTTCCTGTGTCGTGAGAAACAGTGAAGTC 380
QY 1525 TGACCAATGGCCAGACACATGCTGCAACTGTAGTCAAGACTGTCCAGGACAGG -T 1582
DB 379 TGACCAATGGCCAGACACATGCTGCAACTGTAGTCAAGACTGTCCAGGACAGG 320
QY 1583 TTGTGACACAGCCCTACTTTGGGACACAGCTGAGGTGTAAGGGCAGACAAACAGGTGA 1642
DB 319 TTGTGACACAGCCCTACTTTGGGACACAGCTGAGGTGTAAGGGCAGACAAACAGGTGA 260
QY 1643 GGGAGTGTGACACAGAGAGAGTCTTCGCTGACCTCACCCTGAGCCCACTCTTACAGAC 1702
DB 259 GGGAGTGTGACACAGAGAGAGTCTTCGCTGACCTCACCCTGAGCCCACTCTTACAGAC 200
QY 1703 TGGAGAGTGCATGACAGCCACACTGATCCTGTTCAGCAGAGTCTCTGCT -CTGTGCCAGGC 1761
DB 199 TGGAGAGTGCATGACAGCCACACTGATCCTGTTCAGCAGAGTCTCTGCTGTGCCAGGC 140

QY 1762 TCTTGTATTATAGCATGATCAGATGTGTGTCAGACTCTTCTGGGCTGG -AGACCAAGGT 1820
DB 139 TCTTGTATTATAGCATGATCAGATGTGTGTCAGACTCTTCTGGGCTGGAGAACCAAGGT 80
QY 1821 CACTTGTACCTGTCTCTGTGACACAGAGTCTTGAAGGATCTCAGGACGCTCAGCCCA 1880
DB 79 CACTTGTACCTGTCTCTGTGACACAGAGTCTTGAAGGATCTCAGGACGCTCAGCCCA 20
QY 1881 AGC-TTCTACCTGTCTTT 1897
DB 19 AGCCTTCTACCTGTCTTT 2

RESULT 21
AL582265/c 595 bp mRNA linear EST 16-FEB-2001
LOCUS AL582265 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL005YM09 3
DEFINITION prime, mRNA sequence.
ACCESSION AL582265
VERSION AL582265.1 GI:12950078
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 595)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1. 595
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DL005YM09"
/clone_11b="LTI_NFL010_BC2"
/sex="male"
/issue_type="B cells from Burkitt lymphoma"
/note="Vector: PCWVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCWVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 113 a 162 c 165 g 150 t 5 others
ORIGIN

Query Match 25.9%; Score 520.4; DB 9; Length 595;
Best Local Similarity 97.0%; Pred. No. 3e-87;
Matches 578; Conservative 5; Mismatches 7; Indels 6; Gaps 5;

QY 1076 CCCACGCGCGGCTCCAGCTCCAGCATGGTTACTATAC -GAATAACTTTGCTTAGAAG 1134
DB 595 CCCACGCGCGGCTCCAGCTCCAGCATGGTTACTATACGGAATAACTTTGCTTAGAAG 536
QY 1135 TCACACTCCCC -AATTCAGATGTCCCC -AAGAAGATATGCAAAAGGCCCCAGGAAGAGT 1192
DB 535 TCACACTCCCAATCCAGATGTCCCAAGAAAGATATGCAAAAGGCCCCAGGAAGAGT 476
QY 1193 CCCAGCTCTACTGGTGGCCAGAGCTGTGCAGAGAGCCAGATGAGAACTGTTGGTG 1252
DB 475 CCCAGCTCTACTGGTGGCCAGAGCTGTGCAGAGAGCCAGATGAGAACTGTTG -GTG 417
QY 1253 CTTTCCCTATTTTTGTCCGCAATGCATCTTAGCCAGAAACAGCCCAAAAGCCAGGT 1312

```
|||||
Db 416 CCTCCCTATTTTTGCGGAAATGCATCCTAGGCCAGAAACAGCCCAAGAGGCCCAAGT 357
Oy 1313 CAAAGTCCTCTGACAGCAAGATGCTGTAGAGACAGGCTTCATGCGCTGCGCCGGA 1372
Db 356 CAAAGTCCTCTGACAGCAAGATGCTGTAGAGACAGGCTTCATGCGCTGCGCCGGA 297
Oy 1373 CAAAAATTCATCCAGCTACTGACACAGTCATGATCCGCCATTCGCTGTTAAACCCAGAG 1432
Db 296 CAAATTCATCCAGCTACTGACAGTCATGATCCGCCATTCGCTGTTAAACCCAGAG 237
Oy 1433 CCAAGGTTAAAGCAGAGGAGGAGGAGAGCCCTGCTCTCTTCACAGCCCAAGCTGG 1492
Db 236 CCAAGGTTAAAGCAGAGGAGGAGGAGAGCCCTGCTCTCTTCACAGCCCAAGCTGG 177
Oy 1493 ACACCTTCCTGCTGCTGAGACAGTCATGACCAATGGCCAGACATGCTGCGAA 1552
Db 176 ACACCTTCCTGCTGCTGAGAACAGTGTGACCAATGGCCAGACATGCTGCGAA 117
Oy 1553 CTTGTAGTCAGAGACTGTTCACAGCAGG--TTTGTGACAGAGCCCTTCTCGGAGCC 1610
Db 116 CTTGTAGTCAGAGACTGTTCACAGCAGGAGTTTGTGACAGAGCCCTTCTCGGAGCC 57
Oy 1611 AGCCTGAGTGTAGAGGCGACAGCAAGAGTGAAGGTGATGTGACACCCAGAGACT 1666
Db 56 AGCCTGAGTGTAGAGGCAVACAAGAGTGAAGGTGATGTGACACCCAGAGACT 1

RESULT 22
Bt151643 702 bp mRNA linear EST 05-JUL-2001
LOCUS 602916189F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:506423 5',
DEFINITION mRNA sequence.
ACCESSION Bt151643
VERSION Bt151643
KEYWORDS Bt151643.1 GI:14611644
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 702)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11179 row: b column: 24
High quality sequence stop: 694.
Location/Qualifiers
1. 702
/organism="Mus musculus"
/strain="C2BEH II (fetal)"
/db_xref="taxon:10090"
/clone="IMAGE:506423"
/tissue_type="NCI_CGAP_Lu29"
/stem_cell_origin="spontaneous tumor, metastatic to mammary.
Stem cell origin"
/lab_host="DH10B"
/Note="Organ: Lung; Vector: PCMV-SpOrF6; Site_1: Salt;
Site_2: Not; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 186 a 182 c 190 g 144 t

ORIGIN
```

```
Query Match 24.8%; Score 498; DB 10; Length 702;
Best Local Similarity 87.4%; Pred. No. 4, 3e-83;
Matches 568; Conservative 0; Mismatches 80; Indels 2; Gaps 2;
```

```
Oy 72 TGGTGCCCTGGGCGCCCTTGAGTCCAGCATATGCTATCCGCTGCTGTGACATATC 131
Db 41 TGGTCCCTGGGCGCTGAGTGAGCAGCATATGCTATCCGCTGCTGTGACATATC 100
Oy 132 CTCGACCTTCTGATACATCCCGGACGAGCGCCGATCCATGCGGCGACACTTCCA 191
Db 101 CTCGACCTTCTGATACATCCCGGACGAGCGCTGCTGATCCATGCGGCGACACTTCCA 160
Oy 192 CTTGACAGTCCATTAATTCAGTCCCTTGAAGACAGCACAAGTGGACCTGCCACAGTCCG 251
Db 161 TCTGCAATGCCCTAATCCAGTGGTTGAGACAGCACCACAAGTGGACCTGCCACAGTGG 220
Oy 252 AATCCAGGTTGGCAAAAAGAACATTATCATATAGCTCTTCTTGATCTTCCACAGAGGA 311
Db 221 AATCCAGGTTGGCAAAAAGAACATTATTAACAACATTTCTTGGACCTGCCAGAGGA 280
Oy 312 GGAGAAATGCTTGGATGAGATTTCTTAAGAAATGAACCTGACATGTCAGACCCAGCT 371
Db 281 GGAGAAATGCTTGGATGAGATTTCTTAAGAAATGAACCTGACATGTCAGACCCAGCT 339
Oy 372 TTCCAGAAAAGACAGAGAGAAACGAGACAGCTCATCATCAGACCTCGGGAGTAC 431
Db 340 TTCCAGAAAAGACAGAGAGAAACGAGACAGCTCATCATCAGACCTCGGGAGTAC 399
Oy 432 GCTGGAAGAACCAATGCTACTGTGATCTCTGACAGAGGCTTGGGCAAGCCAGAT 491
Db 400 CCTGGAAGAACCAATGCTACTGTGATCTCTGACAGAGGCTTAAACAAGGACAGAT 459
Oy 492 GCTGTGCTCCACACTGAAAAAGAGATGAAGTACTTAAGACAGCAGAGATGAGACAA 551
Db 460 GCTGTGCTCCACACTGAAAAAGAGATGAAGTACTTAAGACAGCAGAGATGAGACAA 519
Oy 552 ACAAGCACAAGAGAGAGCGCGGCTCAGAGCAAGATGAAGACATGAGAGCATGGA 611
Db 520 ACAAGCACAAGAGAGAGCGCGGCTCAGAGCAAGATGAAGATGAAGATGAAGCAAAATTT 579
Oy 612 GCTTCTACTCCAGAGCAGCTCCCTGAGTGAGAGGATATCCGAGCATGGGTGGG 671
Db 580 GCTTCTACTCCAGAGCAGCGGCTGAGTGAGAGGATATTCGAGCATGGGTGGG 639
Oy 672 AC-AGTCAGCGGTGAAGAACAGCTGCTGTACGTGTCTCAAGAA 720
Db 640 ACAAGTCAGCGGTGAAGAACAGCTGCTGTCTCTCAAGAA 689

RESULT 23
AL563489 905 bp mRNA linear EST 16-FEB-2001
LOCUS AL563489/c
DEFINITION prime, mRNA sequence.
ACCESSION AL563489
VERSION AL563489.1 GI:12912932
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 905)
Li, W.B., Gruber, C., Jessup, J., and Polayes, D.
Full-length cDNA libraries and normalization
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 905
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source
```

/clone="CSDD006YH09"
/clone_lib="LTI_NFL001_NBCA"
/sex="male"
/tissue.type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a Noli-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (301) 301 610 8371 Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com"

BASE COUNT 214 a 227 c 276 g 172 t 16 others
ORIGIN

Query Match 24.4%; Score 490.4; DB 9; Length 905;
Best Local Similarity 91.1%; Pred. No. 1,1e-81;
Matches 560; Conservative 4; Mismatches 46; Indels 5; Gaps 4;

QY 1382 TCACGCTACTGACACATGATCGCCCATTCCTGTTAAGCCCAAGCAAGTTA 1441
DB 616 TCCTCCAACTAAAMAGTCATGAMCGCCCATTCCTGTTAAGCCCAAGCAAGTTA 557
QY 1442 AGCAGAGGGTGAAGGTTGAACCGTGCCTCTCTCT-TCCAGGCCAAGCTGGACCTTC 1500
DB 556 AGCAGAGGGTGAAGGTTGAACCGTGCCTCTCTCTCTCCAGGCCAAGCTTC 497
QY 1501 CTGTGTCGTGAGAACAGTGTGACCAATGGCCAGACATGCTGCACACTGTGAGG 1560
DB 496 CTGTGTCGTGAGAACAGTGTGACCAATGGCCAGACATGCTGCACACTGTGAGG 437
QY 1561 TCAGGACTGTCCAGCAGGG--TTTGTGACAGAGCCCTACTTTTGGGACCAAGCTGAG 1618
DB 436 TTCAGGACTGTCCAGCAGGGTTTGTGACAGAGCCCAACTTTTGGGACCAAGCTGAG 377
QY 1619 GTGTAAGGGCAGAACACAGTGTGAGGTGAGTGTGACACCCCAAGACTGCTTTCCTGCC 1678
DB 376 GTGTGAGGGCAGAACACAGTGTGAGGTGAGTGTGACACCCCAAGACTGCTTTCCTGCC 317
QY 1679 CTCACCTGCCCCACTCTCTACGACTGGAGCTGACATGACACGCCCACTGATCCTGTGAG 1738
DB 316 CTCACCTGCCCCACTCTCTACGACTGGAGCTGACATGACACGCCCACTGATCCTGTGAG 257
QY 1739 CAGGTCTCTGCT-CTGTTGCCAGGCTTTGTTATATAGCCATGATGATGATGATGATGATG 1797
DB 256 CAGGTCTCTGCT-CTGTTGCCAGGCTTTGTTATATAGCCATGATGATGATGATGATGATG 197
QY 1798 TTTTGTGGGCTGGAGACAGGTCACCTGTTGATGCTGTGCTGTGGACACAGTCTGTGAG 1857
DB 196 TTTTGTGGGCTGGAGACAGGTCACCTGTTGATGCTGTGCTGTGGACACAGTCTGTGAG 137
QY 1858 GCATCTGAGCAGCCTCAGCCCAAGCTTCTACCTGCTTGTGACTGTGCTTGTAGATAGCC 1917
DB 136 GCTTCTGAGCAGCCTCAGCCCAAGCTTCTCCTGCTTTCACCTGCTTGTAGATAGCC 77
QY 1918 TGGGCCAAGAGGGTGGGGAATGAGATATGACATGAGATGATGAGAGATGAGATGAGAT 1977
DB 76 TGGGCCAAGAGGGTGGGGAATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 18
QY 1978 TTTTCCGAAAAAAA 1992
DB 17 TTTTCTATGTATTAATA 3

RESULT 24
BF162255 744 bp mRNA linear EST 30-OCT-2000
LOCUS BF162255 601770948F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3990371 5',
DEFINITION mRNA sequence.

ACCESSION BF162255
VERSION BF162255.1 GI:11042462
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE NIH-MGC http://img.nci.nih.gov/
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 744)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9201 row: 0 column: 12
High quality sequence stop: 730.

FEATURES
source location/Qualifiers
1..744

/organism="Mus musculus"
/strain="C57BL/6J (f1er1)"
/db_xref="taxon:10090"
/clone="IMAGE:3990371"
/clone_lib="NCI_CGAP_Lu29"
/tissue.type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"

/note="Organ: Lung; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 201 a 183 c 212 g 148 t
ORIGIN

Query Match 24.4%; Score 489.6; DB 10; Length 744;
Best Local Similarity 84.7%; Pred. No. 1,6e-81;
Matches 608; Conservative 0; Mismatches 104; Indels 6; Gaps 5;

QY 26 AGCGAGCGGGCGGCTCTACGAACCGGACCTGTACCACTTCTTGGCTGCGTGGCC 85
DB 16 AGCGAGCGGGCGGCTCTACGAACCGGACCTGTCTGTGCTG-GCAGCTGTTCCCTGGCGT 74
QY 86 CTTGAGTCCAGCAGCATCATGCTATCGCTGTGTCACATATCTGCTCGACTTGTG 145
DB 75 GCTTGAATGAGCAGCATCATGCTATCGCGCTCTGTGCACTATCTGCTCGACTTGTG 134
QY 146 ATCACTCCCGCAGAGTGGCGGCATCCACTGCGGCCACACCTTTCACCTGACGTCTAA 205
DB 135 ATCACTCCCGCAGAGTGGCGGCATCCACTGCGGCCACACCTTTCACCTGACGTCTAA 194
QY 206 TTTGACTCTTTGGAGACGACCAAGTGGAGCTGCGCCAGCAGTGGCCGAATTCAGATGGCA 265
DB 195 TTTGACTCTTTGGAGACGACCAAGTGGAGCTGCGCCAGCAGTGGCCGAATTCAGATGGCA 254
QY 266 AAGAACCATTTATCAATTAAGCTTCTTATGATTTTGGCCAGGAGAGAGATGCTTGG 325
DB 255 AAGAACCATTTATTAACAACTTTTCTTATGACCTGCGCCAGGAGAGAGATGCTTGG 314
QY 326 ATGAGCAATTTTAAAGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 385
DB 315 ATGAGCAATTTTAAAGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 374
QY 386 AGGAGAAAGAGCAGCAGCAGTCAATGACATGACATGACATGACATGACATGACATGACATGAC 445
DB 375 GGGGAAAGAGCAGCAGCAGTCAATGACATGACATGACATGACATGACATGACATGACATGAC 434
QY 446 ATGCTACTGTGTATCTCTGACAGAGGCTTGGGCAAGGCCGAGAGTGTGCTGCCACAC 505

QY	26	AGCGAGGCGGGGGCCCTCTACGAAGCCGGACCTGTAGACGTTTCTTTGGCTGCCGGGCC	85
Db	21	AGCGGTGGCCGGTTTCAGCCACCACCAACTGTGTCGTG-GCAGCTGGTTCCCGGGCT	79
OY	86	CCTTTAGTCACGACCTATCATGCGCTATCCGTGTGTGACTATATGCTCCGACTCTTCG	145
Db	80	GCTTAGTCGAGCCATCATGCTATCCGGCGGTCTGTGACATATCTGCTCCGACTCTTCG	139
OY	146	ATCACTCCCGCAGCTGGCGCCCATCCACTGCGGCCACACCTTCCACTTTGCAGTCCTA	205
Db	140	ATCACTCCCGCAGCTGGCGCCCATCCACTGCGGCCACACCTTCCACTTTGCAGTCCTA	199
OY	206	TTCACTCCCTTGTAGACAGACACCAAGTCGGACCTGCCACAGCGCCAGTTCAGATTCAGTTGGCA	265
Db	200	TCCAGTGGTTTGTAGACAGACACCAAGTCGGACCTGCCACAGTGTGAATTCAGTTGGCA	259
OY	266	AAAAGACATTAATCAATAAAGCTCTTTTGTATCTTTGACCTTGGCCAGAGAGAGAGAAATGCTTTGG	325
Db	260	AAAAGACATTAATTAACAACACTTTCTTTGTACCTGCCAGAGAGAGAGAAATGCTTTGG	319
OY	326	ATCGAGATTCCTTAAGAATGAACAGGACAAATGTCAGAGCCAGCTTTCCAGAAAGACA	385
Db	320	ATCGAGATTCCTTAAGAATGAACAGGACAAATGTCAGAGCTAGCTTTCCAGAAAGACA	379
OY	386	AGGAGAAACGAGACAGCCAGGTCAATCAGCACTCTCGGGATATCGCTGGAAGAACCA	445
Db	380	GGGAGAAACGAGACAGCCAGGTCAATCAGCACTCTCGGGATATCGCTGGAAGAACCA	439
OY	446	ATGCTACTGTGTATCTCTGCAGCAGGCCCTTGGGCAAGGCCGAGATGCTGTGCTCCACAC	505
Db	440	ATGCTACTGTGTATCTCTGCAGCAGGCCCTTGGGCAAGGCCGAGATGCTGTGCTCCACAC	499
OY	506	TGAAAACACAGATGAAGTACTTGAAGCAGCAGCAGATGAGAACCAACACACACAGAGG	565
Db	500	TGAATTAACAGATGAAGTACTTGAAGCAGCAGCAGATGAGAACCAACACACACAGAGG	559
OY	566	AGGCGGGCGCGCTCAGAGCAAGATGAAGACCATGAGCAGCATTTGAGCTTACTCCAGA	625
Db	560	AGGCGGGCGCGCTCAGAGCAAGATGAAGACCATGAGCAGCATTTGAGCTTACTCCAGA	619
OY	626	GCCAGCTCCCTGAGGTGTGAGAGATGATCCGAGCATGGGTGTGGACAGTACAGCGGAGG	685
Db	620	GCCAGCTCCCTGATGTGTGAGAGATGATACAGAAATGTGTGTGACAGTACAGCGGTGG	679
OY	686	AACAGCTG 693	
Db	680	AGCATGTG 687	
RESULT	26		
LOCUS	AA564143/c		
DEFINITION	523 bp mRNA	linear	EST 08-SEP-1997
ACCESSION	AA564143		
VERSION	nk44f02.s1	NCI- <u>CGAP</u> -GC2	Homo sapiens cDNA clone IMAGE:1016379 3'
KEYWORDS	AA564143		
SOURCE	AA564143.1	GI:2335782	
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:		
TITLE	Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.		
JOURNAL	NCI- <u>CGAP</u> http://www.ncbi.nlm.nih.gov/ncicgap .		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index		
	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgaps-remail.nih.gov		
	Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R.		
	Emmert-Buck, M.D., Ph.D.		
	cDNA Library Preparation: Stratagene, Inc., David B. Krizman,		
	Ph.D.		

Db	422	GGGGCGTGAAGTGGAGAGAAATGATCCGAGAAATGGGGTGTGGGACAAATCAGTCAAGTGGAGACA	481
QY	690	GCTGGCTGTGTACTGTGTCTCTCAAGAAAGATGATCAGAAATCTPAAAGAGCAGGAA	749
Db	482	GTGGCTGTGTACTGTGTCTCTCTCAAGAAAGATGATCAGAAATCTPAAAGAGCAGGAA	541
QY	750	GGCTTC 755	
Db	542	GGCTTC 547	
RESULT 28			
LOCUS	BI341090	571 bp	mRNA linear EST 30-JUL-2001
DEFINITION	366355 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.		
ACCESSION	BI341090		
VERSION	BI341090.1	GI:15034379	
KEYWORDS	EST.		
SOURCE	PIG.		
ORGANISM	Sus scrofa		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
AUTHORS	1 (bases 1 to 571) Fahrenkrug,S.C., Fieking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W. and Keele,J.W.		
TITLE	Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smtlhe@mail.marc.usda.gov Single pass sequencing. Bases called and alt. trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options. PCR primers FORWARD: AGGAACAGCTATGACCAT BACKWARD: GTTTCCTCAGTCAAGAG Plate: 107 row: K column: 10 Seq primer: ATTAGTGACATATAG. Location/Qualifiers 1..571 /organism="Sus scrofa" /db_xref="taxon:9823" /clone_id="MARC_2P1G" /tissue_type="pooled" /lab_host="DH10B" /note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI; Library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta."		
FEATURES			
source			
BASE COUNT	154 a 162 c 146 g 109 t		
ORIGIN			
Query Match	22.1%	Score 443:	DB 10: Length 571:
Best Local Similarity	88.3%	Pred. No. 8.1e-73:	
Matches	506; Conservative	0; Mismatches	60; Indels 7; Gaps 2;
QY	11	GGGAATTTGAAGCAGGCGGCGC----	GGGCGCTTACGAGCGGAGCTGTAGCACT 66
Db	1	GGCAATTTGAAGCAGGCGGCGGCGCCTGAGGAGCGGCAGCAAGCTGAGACCCCTTGCGGCT 60	
QY	67	TTCTTTGGCTGCTGGGCGCCCTTGAGTCCAGCATCATGCTCATCCGTGCTGTGCACCT 126	
Db	61	TTCTTTGGCTGCTGGGCGCCCTTGAGTCCAGCATCATGCTCATCCGTGCTGTGCACCT 117	
QY	127	ATTCCTCCGACCTTTTGATGATCATCCGCGGCGAGCGGCGGCGCATCCTACTGGGGCAGACG 186	
Db	118	ATATGCTCCGACCTTTTGATGATCATCCGCGGCGAGCGGCGGCGCATCCTACTGGGGCAGACG 177	

QY	187	TTGACATGCTAGTGGCTATTCAGTCCTTTTGAGACAGACCAAGTCGGACCTGGCCACAG	246
Db	178	TTTCCATCTGCAATGGCTTAATTCAGTGTGTCGAGACAGACCAAGTCGGACCTGGCCACAG	237
QY	247	TGCCCAATTCAGGTTGGGCAAAAAGAACATTTATCAATAAGCTCTTCTTTGATCTTGGCCAG	306
Db	238	TGCCCAATTCAGGTTGGGCAAAAAGAACATTTATCAATAAGCTCTTCTTTGATCTTGGCCAG	297
QY	307	GAGGAGGAGAAATGTCCTTGGATCGAGAATTTCTTAAAGAAATGAACCTGGACAAATGTCAGAGCC	366
Db	298	GAGGAGGAGAGAGTCTTATGATGTCAGAAATTTCTTAAAGAAATGAACCTGGATTAACCAAGAGCC	357
QY	367	CAGCTTTCCCGAAGAAAGACAGAGAAACAGACAGACCCAGTTCATCATGACACTCTGGCG	426
Db	358	CTGCTTTCCCGAAGAAAGACAGAGAAACAGACAGACCCAGTTCATCATGACACTTTGGCG	417
QY	427	GATGACGCTGGAAGAACGCAATGCTCTGCTGATCTGTCGACAGACGCTTGGGCAAGGCC	486
Db	418	GACACGCTGGAAGAACGCAATGCTCTGCTGATCTGTCGACAGACGCTTGGGCAAGGCC	477
QY	487	GAGATGCTGTGCTCCACACTGAAAAAGCAGATGAAGTACTTAGACAGACAGCAGATGAG	546
Db	478	GAAATGCTGTGCTCCACACTGAAAAAGCAGATGAAGTACTTAGACAGACAGCAGATGAG	537
QY	547	ACCAACCAAGCACACAGAGAGAGCGGCGCGCTC	579
Db	538	ACCAACCAAGCACACAGAGAGAGCGGCGCGCTC	570
RESULT	29		
LOCUS	BG235999/c		
DEFINITION	na120109.x1 Soares_NPBMC Homo sapiens CDNA clone IMAGE:4141600 3', mRNA sequence.		
ACCESSION	BG235999		
VERSION	BG235999.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 495)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: M. Bento Soares, Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. and M. Fatima Boraldo, Ph.D. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: infoimage.lnl.gov Seq primer: -40UP from Glbco High quality sequence stop: 416. Location/Qualifiers 1..495 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4141600" /issue_lib="Soares_NPBMC" /tissue_type="lymphocyte" /lab_host="DH10B (phage-resistant)" /note="Organ: blood; Vector: pT73D-Pac; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was primed with a NotI - oligo(dT) primer [5', TGTTACCAATCTCAAGTCAGAGGCGCGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with NotI and cloned into the NotI and Eco RI sites of the modified pT73 vector. Library		
FEATURES			
SOURCE			

is normalized; constructed in the laboratory of M. Bento Soares (University of Iowa)."
BASE COUNT 111 a 137 c 123 g 124 t
ORIGIN

Query Match 21.1%; Score 422.6; DB 10; Length 495;
Best Local Similarity 97.1%; Pred. No. 5.2e-69;
Matches 473; Conservative 0; Mismatches 9; Indels 5; Gaps 4;

```
QY 1525 TGACCAATGGCCAGACACATGCTTGGCACTTGTAAGTCAAGAGACTGTCAGGCGAGGG--T 1562
DB 495 TGACCAATGGCCAGACACATGCTTGGCACTTGTAAGTCAAGAGACTGTCAGGCGAGGGTT 436
QY 1583 TTGTGACACAGGCCCTTCTGGGACACAGCTGAGGTGTAAGGGACAGACAAAGGTGA 1642
DB 435 TTGTGACACAGGCCCTTCTGGGACACAGCTGAGGTGTAAGGGACAGACAAAGGTGA 376
QY 1643 GGGTGAAGTGAACACCCAGAGACTGCTCTGCCCCACACCCCTGCCCCACCTCTACGAC 1702
DB 375 GGGTGAAGTGAAGACCCAGAGACTGCTCTGCCCCACACCCCTGCCCCACCTCTACGAC 316
QY 1703 TGGAGACTGCATGACACGCCCTGATCTCTGTCAGAGGTCTGCT--CTGTTGCCAGGC 1761
DB 315 TGGAGACTGCATGACACGCCCTGATCTCTGTCAGAGGTCTGCTCTGTTGCCAGGC 256
QY 1762 TCTTGTATAGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1821
DB 255 TCTTGTATAGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 196
QY 1822 ACTTGTGACTGCTCTGTCGAGACAGAGCTCTGAGCATCTCAGGACGCTCAGCCCAA 1881
DB 195 ACTTGTGACTGCTCTGTCGAGACAGAGCTCTGAGCATCTCAGGACGCTCAGCCCAA 136
QY 1882 GCTTTACTGCTCTGTCGAGACAGAGCTCTGAGCATCTCAGGACGCTCAGCCCAA 1940
DB 135 GCTTTACTGCTCTGTCGAGACAGAGCTCTGAGCATCTCAGGACGCTCAGCCCAA 76
QY 1941 GAGGATGACATGAGATGATGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 2000
DB 75 GAGGATGACATGAGATGATGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 17
QY 2001 AAAAAA 2007
DB 16 AAAAAA 10
```

RESULT 30
AM248701/c 478 bp mRNA linear EST 07-JAN-2000
LOCUS 2821007.3prlme NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2821007 3',
DEFINITION mRNA sequence.
ACCESSION AM248701.1 GI:6591694
VERSION AM248701.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 478)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Other_ESTS: 2821007.5prlme
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DTF CDNA Library Preparation: Ling
Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E.
Consortium (LNL) DNA Sequencing by: Berkeley MGC sequencing
project Clone distribution: MGC clone distribution information can
be found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bdrip/image/image.html Base Calling / Quality
Scores: PHRED from University of Washington Genome Center. Vector

Trimming: cross-match from University of Washington Genome Center
PHRAP suite. Poly-T identification: patmatch.pl from Berkeley
Drosophila Genome Project. Universtiy of Washington Genome Center:
<http://www.genome.washington.edu> polydenylation: Based upon the
presence of a XhoI site followed by a run of 14 or more T residues
at the beginning of the sequence, this cDNA insert was
polyadenylated.
Plate: LICMS row: K column: 24
High quality sequence stop: 458.

FEATURES
source
location/Qualifiers

```
1..478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2821007"
/clone_1ib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOT87; Site_1: XhoI; Site_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCAAGAG(G). Size selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 111 a 134 c 120 g 112 t 1 others
ORIGIN

```
Query Match 20.4%; Score 410.4; DB 9; Length 478;  
Best Local Similarity 96.5%; Pred. No. 9.9e-67;  
Matches 462; Conservative 0; Mismatches 12; Indels 5; Gaps 4;
```

```
QY 1532 TGCCACAGACACATGCTGACATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1589
DB 478 TGCCACAGACACATGCTGACATTTGATGATGATGATGATGATGATGATGATGATGATGAT 419
QY 1590 CAGAGCCCTACTTCTGGGACACAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1649
DB 418 CAGAGCCCTACTTCTGGGACACAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 359
QY 1650 TGTGACACCCAGAGACTGCTCTTCTGCTCCCTACACCTGCCCCACTCTCTAGACTGGAGC 1709
DB 358 TGTGACACCCAGAGACTGCTCTTCTGCTCCCTACACCTGCCCCACTCTCTAGACTGGAGC 299
QY 1710 TGACATGACACGCCACTGATCTGTGACAGAGTCTGCT--CTGTTGCCAGGCTCTTGT 1768
DB 298 TGACATGACACGCCACTGATCTGTGACAGAGTCTGCTCTGTTGCCAGGCTCTTGT 239
QY 1769 TATAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1828
DB 238 TATAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 179
QY 1829 GACTGTCTCTGTGACACAGAGTCTTGAAGCATCTCAGGACGCTCAGGCCAAGCTTCTA 1888
DB 178 GACTGTCTCTGTGACACAGAGTCTTGAAGCATCTCAGGACGCTCAGGCCAAGCTTCTA 119
QY 1889 CTTGCTTTTGAAGTCTTCTGTA--GCATAGCCTGAGCCAAAGCAGGCTGGGATGAGAGATA 1947
DB 118 CTTGCTTTTGAAGTCTTCTGTA--GCATAGCCTGAGCCAAAGCAGGCTGGGATGAGAGATA 59
QY 1948 GACATGGAGATGATGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 2006
DB 58 G-CATGGAGATGATGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1
```

RESULT 31
BE304189 519 bp mRNA linear EST 13-JUL-2000
LOCUS BE304189
DEFINITION 601086421F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3500621 5',
mRNA sequence.
ACCESSION BE304189

```

VERSION      BE304189.1  GI:9174333
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
REFERENCE    Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
              Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
              1 (bases 1 to 519)
AUTHORS      NIH-MGC http://mgc.ncl.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs.femail.nih.gov
              Tissue Procurement: Jeffrey Green M.D.
              cDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNL at:
              http://image.llnl.gov
              Plate: LL8559 row: 1 column: 06
              High quality sequence stop: 515.

FEATURES
  source
    1..519
      /organism="Mus musculus"
      /strain="FVB/N"
      /db_xref="taxon:10090"
      /clone="IMAGE:3500621"
      /clone_lib="NCI_CGAP_Mam6"
      /sex="female, virgin"
      /tissue_type="infiltrating ductal carcinoma"
      /dev_stage="5 months"
      /lab_host="DH10B"
      /note="Organ: mammary; Vector: pCMV-Sport6; Site:1: SalI;
      Site:2: NotI. Cloned unidirectionally. Primer: oligo dt.
      Library constructed by Life Technologies. Investigator
      providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT   159 a 116 c 152 g 92 t

ORIGIN
Query Match      20.2%: Score 406.2; DB 10; Length 519;
Best Local Similarity 86.8%: Pred. No. 5.9e-66;
Matches 447; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

OY 254 TCCAGTTGGCAAGAACCATTTATCAATAGCTCTCTTGGATGGCCAGAGAGAG 313
      |||||
Db 5 TCCAGTTGGCAAGAACCATTTATTAACAACATTTCTTTGACCTGCCAGAGAGAG 64

OY 314 AGAATGCTTGGATCGAATTTCTTAAGAATGAAGTGGACATGTGAGCCAGCTTT 373
      |||||
Db 65 AGAATGCTTGGATCGAATTTCTTAAGAATGAAGTGGACAGCTCAAGCTCAGCTTT 124

OY 374 CCCGAAAGACAGAGAAAGAGACAGCCAGCTCATCTGACACATCTGCGGATGCGC 433
      |||||
Db 125 CCCGAAAGACAGAGAAAGAGAGACAGCCAGCTCATCTGACACATCTGCGGAGACACC 184

OY 434 TGAAGAAGCAATGCTACTGTGATCTCTGACAGAGCCCTTGGCAAGGCCGATGTC 493
      |||||
Db 185 TGAAGAAGCAATGCTACTGTGATCTCTGACAGAGCCCTTAAACAAGCAGAGATGC 244

OY 494 TGTGCTCCACACTGAAAAAGCATGAAGTATTAGACAGACAGCAGATGAGACCAAC 553
      |||||
Db 245 TGTGTTCCACCTGAAAAAGCATGAAGTTCCTGAGACAGCCGAGATGAGACCAAC 304

OY 554 AAGCACAAGAGAGAGCGCGGCTCAGAGAGCAAGATGAGACCATGAGACATTTGAGC 613
      |||||
Db 305 AAGCTCGGAGAGAGCGCGGCTCAGAGAGCAAGATGAGACCATGAGACATTTGAGC 364

OY 614 TTCTACTCCAGAGCAGCTCCCTGAGGTGAGAGAGATGATCCGAGACATGGGTGGGAC 673
      |||||
Db 365 TTCTACTCCAGAGCAGCTCCCTGAGGTGAGAGAGATGATCCGAGACATGGGTGGGAC 424

OY 674 AGTCAGCGGTGGAACAGCTGGCTGTGATCTGTGCTCTCAAGAAGAGTACGAAATC 733

```

```

|||||
Db 425 AGTCAGCGGTGAGACAGCTGGCTGTGATCTGCTGCTCCCTCAAGAAGATGAGAAATC 484

OY 734 TAAAGAGCGCAGAGAGCGCTCAGCGGAGGTGCT 768
      |||||
Db 485 TGAAGAGAGCTCGAGAGCGCACAGCGGAGAACTGGCT 519

RESULT 32
AA684194 554 bp mRNA linear EST 09-DEC-1997
LOCUS     AA684194
DEFINITION
  YN68410..s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
  IMAGE:1003411 5' similar to TR:008854 008854 MTRIP. ;, mRNA
  sequence.
ACCESSION AA684194
VERSION    AA684194.1  GI:2670780
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
              Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
              1 (bases 1 to 554)
AUTHORS    Maria M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
              Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
              Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
              Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
              Meteston, R.
TITLE       The WashU-HMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT     Contact: Maria M/Mouse EST Project
              WashU-HMI Mouse EST Project
              Washington University School of Medicine
              444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@wustl.edu
              This clone is available royalty-free through LNL; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:567627
              Possible reversed clone: similarity on wrong strand
              High quality sequence stop: 512.

FEATURES
  source
    1..554
      /organism="Mus musculus"
      /strain="B6D2 F1/J"
      /db_xref="taxon:10090"
      /clone="IMAGE:1003411"
      /clone_lib="Knowles Solter mouse 2 cell"
      /tissue_type="embryo"
      /dev_stage="2-cell"
      /lab_host="DH10B"
      /note="Organ: embryo; Vector: pBluescribe (modified);
      Site_1: MuuI; Site_2: SalI; Cloned unidirectionally from
      mRNA prepared from 13,500 2-cell stage embryos. Primer:
      SalI(dT): 5'-CGGTGCGACGCTGACGCGTGTGTGTGTGT-3'.
      CDNAS
      were cloned into the MuuI/SalI sites of a modified
      pBluescribe vector using commercial linkers (NEB).
      Average insert size: 1.2 kb."

BASE COUNT   149 a 151 c 139 g 115 t

ORIGIN
Query Match      19.9%: Score 398.6; DB 9; Length 554;
Best Local Similarity 86.6%: Pred. No. 1.5e-64;
Matches 451; Conservative 0; Mismatches 69; Indels 1; Gaps 1;

OY 72 TGGCTGCTCCCTTGGAGTCCAGCAGCATGCTATCCGTGCTCTGTGCACTATCTG 131
      |||||
Db 34 TGGTTCCTCGGCGCTGTGAGTGCAGCAGCATGCTATCCGCGCTCTGTGCACTATCTG 93

OY 132 CTCGACTTCTTGATCAGTCCCGGAGCGTGGCCGATCCATCGGCGGCAACACTTTCA 191
      |||||
Db 94 CTCGACTTCTTGATCAGTCCCGTGAAGCGTGGCGCATCCACTGATGCGCAACTTTTCA 153

```

OY		192	CTTGACATGCCCTAAATTGACTCCTTTTGAAGACAGCAACAGTCGTGGACCTCCACAGTGC	251
Dd		154	TCTGCATAATGCTTAATCCAGTGTGTTGAGCACACACCAAGTGGACCTGCACAGGTAG	213
OY		252	AATCCAGGTGTGGCAAAGAACCATTATTCATTAAGCTCTTCTTGAATCTTGTCCCGAGAGA	311
Dd		214	AATCCAGGTGTGGCAAAGAACCATTATTAACAACACTTTCTTTGACCTGTGCCAGGAAGA	273
OY		312	GGAGAAATGCTTTGGATTCAGAAATTTCTTAAAGATGAACTGGACAATGTCAAGCCCAGCT	371
Dd		274	GGAGAAATGCTTTGGATTCAGAAATTTCTTAAAGATGAACTGGACAACGCTCAAAAGCTCAGCT	333
OY		372	TTCCCCAAGAAGCAAGAGAAACGAGAACGAGCCAGGCTCATCTTCGCACTTCGCGGGATAC	431
Dd		334	TTCCCCAAGAAGCAAGAGAGAAACGAGAACGAGCCACTTTATCGCACTTCTACGGGACAC	393
OY		432	GCTGGAGAAGCAAGCAATGCTACTGTGTGATCTCTCGACGAGGCGTTGGCAGGCGGAGAT	491
Dd		394	CCTGGAGAAGCAAGCAATGCTACTGTGTGATCTCTCGACGAGGCGTTGAACAGGCGAGAT	453
OY		492	GCTGTGCTCCACACT-GAAAAAGCAGATGAAGTACTTGAAGCAGCAGCAGATGAGACCA	550
Dd		454	GCTGTGCTCCACACTCGGAAAAAACGATGAAGTCTTGACACGCGGAGGATGAGACCA	513
OY		551	AACAGCACAAAGAGGAGGCGGCGCGCTCAGAGCAAGATG	591
Dd		514	AACAAGCTCGGAGAGGAGGCCACCGCACTCAAGTGAATG	554
RESULT	33			
BE334637				
LOCUS		589 bp	mRNA	linear EST 14-JUL-2000
DEFINITION		usrl1912.y1 NCI CGAP Mam4 Mus musculus CDNA clone IMAGE:3257638 5'		
		similar to TR:O08854 O08854 TRAF-INTERACTING PROTEIN ;, mRNA		
SEQUENCE.				
ACCESSION		BE334637		
VERSION		BE334637.1 GI:9208413		
KEYWORDS		EST.		
SOURCE		house mouse.		
ORGANISM		Mus musculus		
		Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
		1 (bases 1 to 589)		
REFERENCE		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
AUTHORS		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
TITLE		Tumor Gene Index		
JOURNAL		Unpublished (1997)		
COMMENT		Contact: Robert Strausberg, Ph.D. Email: cga@bsf-research.nhlbi.nih.gov Tissue procurement: Lothar Hennighausen Ph.D., Priscilla Furch Ph.D. CDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: image.llnl.gov/image/html/lresources.shtml		
FEATURES				
source		MGI:1069002		
		Seq primer: -40RP from Glbcoc		
		High quality sequence stop: 362.		
		Location/Qualifiers		
		1..589		
		/organism="Mus musculus"		
		/strain="NMRI"		
		/db_xref="taxon:10090"		
		/clone_image="IMAGE:3257638"		
		/clone_lib="NCI_CGAP_Mam4"		
		/tissue_type="tumor, gross tissue"		
		/dev_stage="5 months"		
		/lab_host="DH10B"		
		/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;		

	Query Match	19.3%, Score 387.4; DB 10; Length 589;
	Best Local Similarity 79.8%; Pred. No. 1,9e-62;	
	Matches 470; Conservative 0; Mismatches 116; Indels 3; Gaps 1;	
OY	672 ACAGTCACCGGTGGGAACAAGCTGCTGTACTGTGTCTCTCAAGAAAGATCAGAGAA 731	
Db	1 AAGTGCACCGGTGGGAACAAGCTGCTGTACTGTGTCTCTCAAGAAAGATCAGAGAA 60	
OY	732 TCTAAGAAGCCAGCACGAGGCCCTCATGGGAGAGTGCCCTACAAGCTGAGAGAGATTGT 791	
Db	61 TCTGAAGAGAGCTCGGAAGGCCACAGGGGAACTGGCTCAGAGTTGAAGAAAGATTGGT 120	
OY	792 TTCCTCCAGAACAAAGTTGCACAGACAGTCTACTCTGAATTGATCATGAGCCCAAGTTAGAACT 851	
Db	121 GTCTCTTAGAGACAGATTGAAGACTCTCAACACTGAGCTGATCAGGCCCAAGTTAGAACT 180	
OY	852 GAAGTCACGCCCAAGAAAGACTTACAGAGTCTGTACACAGAAATCATGACCCTGAAAAAGAA 911	
Db	181 GAGGTTCAGCCCAGAAAGACTTACAAAAGTGTGACACAGAGATCACAGACCTTAAGMAAAM 240	
OY	912 GCTAACGATGTCTGCAGAAACCTTGAACCTGCGACAGGCGCAGTGTGAGCTGTGCAGC 971	
Db	241 GCTGATGATCTCTCCAGGAACTTBAAGCTGTGACCGACCATGTAGAGCGGTAGCCG 300	
OY	972 CCTGGTTTTAGAGAGCCCCAGCCCTGTGGA--GSGTGAATCTGAAGCTCCGCGGCCATC 1028	
Db	301 CTTGGTTTTTAGAGAGCCCCAGCCCTGTGGAAGATGATGAACCCGAGGCTTACACAGCCACC 360	
OY	1028 CTTCGGTGATGATTATGATCTCAATGCTAATCTTTATGTGATFACCTCCCCAGCCGCC 1088	
Db	361 CTTCGGTGATGATGATTATCTCAATAACCAACCTGTATGTAATACCCCTGCMAACCCAAGAC 420	
OY	1089 CTCACGATCCCGACGATNGTTACTACGAAAAACTTTGGCCTAGAGAGTCAACATCCCCCAT 1148	
Db	421 CTCCTGCTCCACGATTTGGCTTCCCAAGAGAGTGTGGCTGGGAGGAGCAGCTCTCCCAT 480	
OY	1149 TCAGAGATGCCCCCAAGAAAGATATGCAAAAGGCCCCAGAAAGAGTCCAGACTCTCACTGGG 1208	
Db	481 GCAGAGATGTCTTCAAAGAAAGTGCAAAAATCTCTCCAAAGCCGAGGTCTCAAGCTTAACTGGG 540	
OY	1209 TGCGCAGAGCTGTGCAGAGAGCCAGATGAGAGAACTGTTGGTGCCTTC 1257	
Db	541 TGCGCAGAGCTGTGTAGAGAGACTGATGATGAGAACTGGCTGTGGCCCTTC 589	
RESULT_34	AII47333 435 bp mRNA linear EST 29-SEP-1998	
LOCUS	AII47333/C gq69g01.s1 Soares_testis_MHT Homo sapiens cDNA clone IMAGE:1839888	
DEFINITION	3', mRNA sequence.	
ACCESSION	AII47333	
VERSION	AII47333.1 GI:3675015	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
JOURNAL COMMENT	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. NCI-CGP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail@nih.gov	

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrrp/image/image.html
Seq primer: -40ml3 fwd. E7 from Amersham
High quality sequence stop: 425.

FEATURES

SOURCE

Location/Qualifiers
1..435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1839888"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGTGGAGGCGGCCCAATTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT773 vector. Library
went through one round of normalization to Cots, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 103 a 124 c 112 g 95 t 1 others
ORIGIN

Query Match 18.6%; Score 373.2; DB 9; Length 435;
Best Local Similarity 97.0%; Pred. No. 8,7e-60;
Matches 423; Conservative 0; Mismatches 8; Indels 5; Gaps 4;

QY 1568 CTGTCCAGGCAAGG--TTTGTGACAGAGCCCTACTTTCGGGACAGCCTGAGTGTAA 1625
|||||
DB 435 CTGTCCAGGCAAGGTTTGTGACAGAGCCCACTTTCGGGACAGCCTGAGTGTAA 376
|||||
QY 1626 GGCAGACAAACAGGTGAGGTGAGTGACACCCAGACAGTCTTCCCTCCACCC 1665
|||||
DB 375 GGCAGACAAACAGGTGAGGTGAGTGACACCCAGACAGTCTTCCCTCCACCC 316
|||||
QY 1686 TGCCCACTCTTACGACTGAGGTGACATACACCCCACTGATCCGTGACAGAGTCC 1745
|||||
DB 315 TGCCCACTCTTACGACTGAGGTGACATACACCCCACTGATCCGTGACAGAGTCC 256
|||||
QY 1746 TGCT-CTGTTCAGAGCTCTTGTATTATGACATGATGATGATGATCTTTCTGG 1804
|||||
DB 255 TGCTCTGTTCAGAGCTCTTGTATTATGACATGATGATGATGATGATCTTTCTGG 196
|||||
QY 1805 GCGTGGAGACAGGTCACCTGTTGACGTCTGTGTGACACAGATGCTTATAGGATCTC 1864
|||||
DB 195 GCGTGGAGACAGGTCACCTGTTGACGTCTGTGTGACACAGATGCTTATAGGATCTC 136
|||||
QY 1865 AGCAGGCTCAGCCCAAGCTTCTACCTGCTTGTACTGCTTCT-AGCATAGCCTGGGCC 1923
|||||
DB 135 AGCAGGCTCAGCCCAAGCTTCTACCTGCTTGTACTGCTTCTTTCAGATAGCCTGGGCC 76
|||||
QY 1924 AAGCAGGTGGGGAATGAGATGACATGAGGATGATGAGAGAGGAAGATTTTCCC 1983
|||||
DB 75 AAGCAGGTGGGGAATGAGATGAGTATGAGAGATGATGAGAGATTTTCAT 17
|||||
QY 1984 GAAAAAATTTTAAAA 1999
|||||
DB 16 GTAAAAATTTAAAA 1

RESULT 35
AM207517/c 455 bp mRNA linear EST 02-DEC-1999
LOCUS
DEFINITION UT-H-B11-afj-f-12-0-UT.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
IMAGE:2721959 3', mRNA sequence.

ACCESSION AM207517
VERSION AM207517.1 GI:6507013
KEYWORDS EST
SOURCE human
ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
1 (bases 1 to 455)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
COMMENT Email: gcaps-remail.nih.gov

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. CDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrrp/image.html The following repetitive
elements were found in this cDNA sequence: 1-29,
>AT-rich#Low-complexity
Seq primer: M13 Forward
POLYA=yes.

FEATURES

SOURCE

Location/Qualifiers
1..455
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2721959"
/clone_lib="NCI_CGAP_Sub3"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NCI_CGAP_Sub3 library is a subtracted library derived from
the NCI_CGAP_Sub1 library, which is a subtracted library
derived from BI. BI constitutes a mixture of 21
normalized or subtracted NCI_CGAP libraries: NCI_CGAP_C04
, NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_C010,
NCI_CGAP_C016, NCI_CGAP_Kid5, NCI_CGAP_Kid12,
NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2,
NCI_CGAP_Br2, NCI_CGAP_C08, NCI_CGAP_CEL1, NCI_CGAP_L012,
NCI_CGAP_Br23, NCI_CGAP_Lu5, NCI_CGAP_Lu24,
NCI_CGAP_Lu19, NCI_CGAP_G04, NCI_CGAP_G06,
NCI_CGAP_Br25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911, 1456008-1456775
, 1500552-1502855); NCI_CGAP_Kid5 pool 1 LLAM 3338-3342
, 3722-3725, 3776-3778 (IMAGE Clonoids 1322912-1325831,
1471368-1472803, 1492104-1493253); NCI_CGAP_Lu5 pool 1
LLAM 3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991,
1520904-1522439); NCI_CGAP_G04 pool 1 LLAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631,
1469064-1470983, 1475582-1476433); NCI_CGAP_Pr22 pool 1
LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids
985608-986759, 1101192-1101959, 121928-1220615);
NCI_CGAP_C010 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE
Clonoids 1057416-1061255, 1144584-1145351). Subtraction
was performed as previously described (Bonaldo, Lennon &
Soares (1996): Normalization and Subtraction: Two
Approaches To Facilitate Gene Discovery. Genome Research
6, 791-806.
TAG_LIB=NCI_CGAP_C04
TAG_TISSUE=colon
TAG_SEQ=CTTCG"

BASE COUNT 103 a 129 c 113 g 108 t 2 others
ORIGIN
Query Match 18.2%; Score 365; DB 9; Length 455;

```
Best Local Similarity 93.3%; Pred. No. 2.9e-58;
Matches 402; Conservative 0; Mismatches 27; Indels 2; Gaps 2;

QY 1579 GGGTTTGGACGAGCCCTACTTTGGGGGACCGAGCTGAGGTGTAAAGGCGAGCAAAACAG 1638
Db 437 GGGTTTGGACGAGCCCTACTTTGGGGGACCGAGCTGAGGTGTAAAGGCGAGCAAAACAG 378
QY 1639 GTGAGGGTGTAGTGTGACACCCAGAGACTGCTCTTCCCTGACCCCTGACCTCTCTTA 1698
Db 377 GTGAGGGTGTAGTGTGACACCCAGAGACTGCTCTTCCCTGACCCCTGACCTCTCTTA 318
QY 1699 CGACTGGAGCTGACATGACACCCAGAGCTGATCTCTGACAGAGTCTCTGCT-CTGTTGCC 1757
Db 317 CGACTGGAGCTGACATGACACCCAGAGCTGATCTCTGACAGAGTCTCTGCTCTGTTGCC 258
QY 1758 AGGCTCTTGTTTATGACCATGATGATGCTGACAGCTTTCTGTGGCTGTGAGACAC 1817
Db 257 AGGCTCTTGTTTATGACCATGATGATGCTGACAGCTTTCTGTGGCTGTGAGACAC 198
QY 1818 GGTCACTTGTGTGACTCTCTGTGACACAGAGTGTGAGGATCTGACAGCAGCCTCAGC 1877
Db 197 GGTCACTTGTGTGACTCTCTGTGACACAGAGTGTGAGGATCTGACAGCAGCCTCAGC 138
QY 1878 CCAAGCTTACCTGCTGCTTGTGACTCTCTTA-GCATAGCCTGGCCCAAGCAGGTGGG 1936
Db 137 CCAAGCTTACCTGCTGCTTGTGACTCTCTTAAGCATAGCCTGGCCCAAGCAGGTGGG 78
QY 1937 AATGAGAGATGACATGGGATGTATGAGAGATGGAAGTTTCCCGAAAAA 1996
Db 77 GAATGAGATGACATGGGATGTATGAGAGATGGAAGTTTCAATTAATTAANTT 18
QY 1997 AAAAAAAAAA 2007
Db 17 AAAAAAAAAA 7

RESULT 36
LOCUS A1695169 399 bp mRNA linear EST 18-DEC-1999
DEFINITION we47h01.x1 NCL_CGAP_Co3 Homo sapiens CDNA clone IMAGE:2344273 3',
mrna sequence.
ACCESSION A1695169
VERSION A1695169.1 GI:4983069
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 399)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert length: 333 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES
source
1. .399
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2344273"
/clone_lib="NCL_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
```

```
/lab_host="DH10B"
/Note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library went through one round of
normalization."
BASE COUNT 93 a 114 c 102 g 90 t
ORIGIN
Query Match 17.5%; Score 351.2; DB 9; Length 399;
Best Local Similarity 97.2%; Pred. No. 1.1e-55;
Matches 389; Conservative 0; Mismatches 8; Indels 3; Gaps 3;

QY 1604 CGGACACGAGCTGAGGTGTAAAGGCGAGACAAACAGGTGAGGTGTGACACCCAGAG 1663
Db 399 CGGACACGAGCTGAGGTGTAAAGGCGAGACAAACAGGTGAGGTGTGAGGTGACACCCAGAG 340
QY 1664 ACTGCTCTTCTCTCCCTACCCCTGCCCCACTCTTACGACTGGAGCTGACATGACAGCC 1723
Db 339 ACTGCTCTTCTCTCCCTACCCCTGCCCCACTCTTACGACTGGAGCTGACATGACAGCC 280
QY 1724 CACTGATTCCTGACAGAGGCTCTGCT-CTGTTGCCAGGCTCTGTTTATAGCATGATCA 1782
Db 279 CACTGATTCCTGACAGAGGCTCTGCTCTGTTGCCAGGCTCTGTTTATAGCATGATCA 220
QY 1783 GATGTGGTCACTCTTCTGAGGCTGTGAGACACGAGTCACTTGTGACTGTCTGTGG 1842
Db 219 GATGTGGTCACTCTTCTGAGGCTGTGAGACACGAGTCACTTGTGACTGTCTGTGG 160
QY 1843 ACCAGAGTGTGAGGATCTCAGGACGCTCAGCCCAAGCTTCTACCTTCTTACTT 1902
Db 159 ACCAGAGTGTGAGGATCTCAGGACGCTCAGCCCAAGCTTCTACCTTCTTACTT 100
QY 1903 GCTTCTA-GCATAGCCTGGGCCAAGCAGGCTGGGAATGAGATGAGATGATAT 1961
Db 99 GCTTCTAAGCATATGCTTGGGCCAAGCAGGCTGGGAATGAGATGAGATGATAT 41
QY 1962 GGAGAGATGGAAGATTTTCCCGAAAAA 2001
Db 40 GGAGAGATGGAAGATTTTCAATTAATTAATAA 1

RESULT 37
LOCUS BE234038 457 bp mRNA linear EST 10-JUL-2000
DEFINITION BE234038 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE234038
VERSION BE234038.1 GI:9018765
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
REFERENCE 1 (bases 1 to 457)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.,
and Keeler,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross-match with the -minscore 18
and -mismatch 12 options.
```

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 77 row: 0 column: 14
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers

FEATURES

source

1..457
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

BASE COUNT 118 a 130 c 115 g 94 t
ORIGIN

Query Match 17.2%; Score 346; DB 9; Length 457;
Best Local Similarity 87.6%; Pred. No. 1e-54;
Matches 403; Conservative 0; Mismatches 50; Indels 7; Gaps 2;

OY 10 AGCGAATTGGAAGCAAGCGAGCG---GGCGCTCTACGAGCCGACCTGTAGCAG 65
DB 1 AGCGAATTGGAAGCAAGCGAGCGAGCGCGCTGAGCGAGCGCGCACTAAGCTGACCTTCGCG 60
OY 66 TTTCTTTGGCTGGCGCCCTTTGAGTCCAGCCATCATGCTATCGCTGCTGTGCGAC 125
DB 61 GTTCTCTGGTACCTGGCCACTGTGAGACAG--TCATGCTATTTGCTCTGTGTGAC 117
OY 126 TATCTGCTCCAGCTTCTTCGATCACTCCGCGAGCGGCGCATCCATCGCGCGCAC 185
DB 118 TATATGCTCCGACTTCTTCGATCACTCCGCGAGCGGCGCATCCATCGCGCGCAC 177
OY 186 CTTTCACCTTGCAGTGCCTAATTCAGTCTTTGAGACGACACCAAGTGGACCTGCCACA 245
DB 178 CTTTCACCTTGCAGTGCCTAATTCAGGAGTTCGAGACGACCAAGTGGAGCTGCCACA 237
OY 246 GTGCCGAAATCAGGTTGGCAAAAGAACATTATCAATTAAGCTCTTTGATGTGGCCA 305
DB 238 GTGCCGAAATCAGGTTGGCAAAAGAACATTATCAATTAAGCTCTTTGATGTGGCCA 297
OY 306 GGAAGAGAGAAATGCTTGGATCGAATTTCTTAAGAAATGAATGCAATGTCAAGAGC 365
DB 298 GGAAGAGAGAGATGCTTGGATCGAATTTCTTAAGAAATGAATGCAATGTCAAGAGC 357
OY 366 CCAGCTTTCCAGAAAGACAGAGAAAGACAGACAGCAGCTATCATTCAGACTGCGG 425
DB 358 CCGCTTTTCCAGAAAGAGAAAGAAAGACAGACAGCAGATCATTCATGACACTTGGCG 417
OY 426 GGATACGCTGGAAAGAGCAATGTCTGATCTCTG 465
DB 418 GGACACGCTGGAAAGAGCGCAACGCCACTGTGAATCCTTG 457

RESULT 38
BM462511 1015 bp mRNA linear EST 05-FEB-2002
LOCUS BM462511
DEFINITION AGNCOUNT_6426349 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5518193
5', mRNA sequence.

ACCESSION BM462511
VERSION BM462511.1 GI:18511551
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1015)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabp-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM2178 row: 3 column: 18
High quality sequence stop: 644.

FEATURES

source

1..1015
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:5518193"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."

BASE COUNT 251 a 278 c 278 g 207 t 1 others
ORIGIN

Query Match 17.2%; Score 345.4; DB 10; Length 1015;
Best Local Similarity 96.1%; Pred. No. 1.2e-54;
Matches 366; Conservative 0; Mismatches 11; Indels 4; Gaps 1;

OY 6 GTGGAGCAAAATTTGAAGCAAGCGAGCGGCG---CTTACGAGAGCGGACCTGTA 61
DB 8 GTCCGCGCAAAATTTGAAGCAAGCGAGCGGCGGCGGCTTACGAGAGCGGACCTGTA 67
OY 62 GCAGTTCTTTGGCTGCTCGGCGCCCTTGAGTCCAGCCATCATGCTATCCGCTCTGT 121
DB 68 GCAGTTCTTTGGCTGCTCGGCGCCCTTGAGTCCAGCCATCATGCTATCCGCTCTGT 127
OY 122 GCAGTTCTGCTCGGCTGCTGATCACTCCGCGAGCGGCGGCGGATCCAGTGGCGC 181
DB 128 GCAGTTCTGCTCGGCTGCTGATCACTCCGCGAGCGGCGGCGGATCCAGTGGCGC 187
OY 182 ACACCTTCCAGTTCAGTGCCTAATTCAGTCTTTGAGACAGCACCAAGTGGAGCTGCC 241
DB 188 ACACCTTCCAGTTCAGTGCCTAATTCAGTCTTTGAGACAGCACCAAGTGGAGCTGCC 247
OY 242 CACAGTGGCGAATCCAGGTTGGCAAAAGAACATTATCAATTAAGCTCTTTGATCTTG 301
DB 248 CACAGTGGCGAATCCAGGTTGGCAAAAGAACATTATCAATTAAGCTCTTTGATCTTG 307
OY 302 CCCAGAGAGAGAGATGCTTGGATCGAGAAATTTTAAAGATGAATGCAATGTCA 361
DB 308 CCCAGAGAGAGAGATGCTTGGATCGAGAAATTTTAAAGATGAATGCAATGTCA 367
OY 362 GAGCCAGCTTTCCAGAAAG 382
DB 368 GAGCCAGCTTTCCAGAAAG 388

RESULT 39
AI633797 379 bp mRNA linear EST 14-DEC-1999
LOCUS AI633797/C
DEFINITION tt28f05.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2242113 3',
mRNA sequence.

ACCESSION AI633797
VERSION AI633797.1 GI:4685127
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 379)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap/
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbirp/image/image.html
Insert Length: 314 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES

source

1..379
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2242113"
/clone_lib="NCI CGAP GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN

92 a 107 c 97 g 83 t

Query Match 16.6%; Score 334; DB 9; Length 379;
Best Local Similarity 97.9%; Pred. No. 1.8e-52;
Matches 370; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

QY 1619 GTGTAAGGCGACAAACAGCTGAGGTGAGTGACACCCAGAGACTGCTTTCTGCC 1678
DB 379 GTGTAAGGCGACAAACAGCTGAGGTGAGTGACACCCAGAGACTGCTTTCTGCC 320
QY 1679 CTCACCTGCCCCACCTCTAGAGCTGGAGCTGACATGACAGCCCACTGATCCTGTCAG 1738
DB 319 CTCACCTGCCCCACCTCTAGAGCTGGAGCTGACATGACAGCCCACTGATCCTGTCAG 260
QY 1739 CAGGTCTGCT-CTGTTGCCAGGCTCTGTTATAGCCATGATCAGATGTCAGACTC 1797
DB 259 CAGGTCTGCTCTGTTGCCAGGCTCTGTTATAGCCATGATCAGATGTCAGACTC 200
QY 1798 TTTCTGGGCGCTGGAGACAGGTCACCTTGTGACTGTCCTGCGACCAAGTCTTGAG 1857
DB 199 TTTCTGGGCGCTGGAGACAGGTCACCTTGTGACTGTCCTGCGACCAAGTCTTGAG 140
QY 1858 GCATCTCAGGCGAGCTCAGGCCAAGCTTCTACCTGCTTGACTTGGACTGGTCTTA-GCATAGC 1916
DB 139 GCATCTCAGGCGAGCTCAGGCCAAGCTTCTACCTGCTTGACTTGGACTGGTCTTA-GCATAGC 80
QY 1917 CTGGGCCAAGCAGGCTGGGGAATGAGAGATAGACATGGATGTATGAGAGATGAGAGA 1976
DB 79 CTGGGCCAAGCAGGCTGGGGAATGAGAGATAG-CAATGGATGTATGAGAGATGAGAGA 21
QY 1977 TTTTCCGAAAAAAA 1994
DB 20 TTTTCATGTAATAAATAA 3

RESULT 40
LOCUS AM002888/c 379 bp mRNA linear EST 08-MAR-2000
DEFINITION wT02c11.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2480372 3',

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

mRNA sequence.
AM002888
AM002888.1 GI:5849804
EST.
human.
Homo sapiens

REFERENCE

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 379)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS

TITLE

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbirp/image/image.html
Insert Length: 482 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES

source

1..379
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2480372"
/clone_lib="NCI CGAP GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN

92 a 107 c 97 g 83 t

Query Match 16.6%; Score 334; DB 9; Length 379;
Best Local Similarity 97.9%; Pred. No. 1.8e-52;
Matches 370; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

QY 1619 GTGTAAGGCGACAAACAGCTGAGGTGAGTGACACCCAGAGACTGCTTTCTGCC 1678
DB 379 GTGTAAGGCGACAAACAGCTGAGGTGAGTGACACCCAGAGACTGCTTTCTGCC 320
QY 1679 CTCACCTGCCCCACCTCTAGAGCTGGAGCTGACATGACAGCCCACTGATCCTGTCAG 1738
DB 319 CTCACCTGCCCCACCTCTAGAGCTGGAGCTGACATGACAGCCCACTGATCCTGTCAG 260
QY 1739 CAGGTCTGCT-CTGTTGCCAGGCTCTGTTATAGCCATGATCAGATGTCAGACTC 1797
DB 259 CAGGTCTGCTCTGTTGCCAGGCTCTGTTATAGCCATGATCAGATGTCAGACTC 200
QY 1798 TTTCTGGGCGCTGGAGACAGGTCACCTTGTGACTGTCCTGCGACCAAGTCTTGAG 1857
DB 199 TTTCTGGGCGCTGGAGACAGGTCACCTTGTGACTGTCCTGCGACCAAGTCTTGAG 140
QY 1858 GCATCTCAGGCGAGCTCAGGCCAAGCTTCTACCTGCTTGACTTGGACTGGTCTTA-GCATAGC 1916
DB 139 GCATCTCAGGCGAGCTCAGGCCAAGCTTCTACCTGCTTGACTTGGACTGGTCTTA-GCATAGC 80
QY 1917 CTGGGCCAAGCAGGCTGGGGAATGAGAGATAGACATGGATGTATGAGAGATGAGAGA 1976
|||||

Db 79 CTGGGCCAAGCAGGGTGGGAATGAGGATAG-CATGGATGTAATGAGAGGATGGAAGA 21
Qy 1977 TTTCCCGAATAAAAAA 1994
||||| | ||| |||
Db 20 TTTTCATGTAAATATAA 3

Search completed: September 4, 2002, 11:29:22
Job time: 10277 sec